Bandwidth Efficiency and Power Efficiency Issues for Wireless Transmissions

A Dissertation
Presented to
The Academic Faculty

by

Ning Chen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
May 2006

Bandwidth Efficiency and Power Efficiency Issues for Wireless Transmissions

Approved by:

Professor G. Tong Zhou, Chair
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Douglas B. Williams
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor J. Stevenson Kenney
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Yang Wang
School of Mathematics
Georgia Institute of Technology

Professor Ye (Geoffrey) Li
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: March 27, 2006

To my maternal grandmother.

ACKNOWLEDGEMENTS

This dissertation drew a period of academic pursuits during which I have been accompanied and supported by many people. It is a pleasure that I now have the opportunity to express my gratitude to all of them.

First and foremost, I would like to express my sincere appreciation and gratitude to my advisor, Dr. G. Tong Zhou, for offering me the opportunity to pursue my Ph.D. degree at Georgia Tech and for entrusting me with this challenging project. She taught me how to strive for excellence in my academic pursuits and helped me to develop critical thinking and presenting skills.

I would like to thank my dissertation committee members: Dr. J. Stevenson Kenney, Dr. Ye (Geoffrey) Li, Dr. Douglas B. Williams and Dr. Yang Wang for taking the time to serve on my committee and for their helpful suggestions. I am also in debt to my M.S. advisor, Dr. Philip De Leon at the New Mexico State University for his warmest encouragement during my graduate studies in the U.S.

I would also like to thank my group members: Lei Ding, Raviv Raich, Hua Qian, Chunpeng Xiao, Thao Tran, Vince Emanuele, Bob Baxley, Chunming Zhao and Kun Shi for the insightful discussions and help along the way. I also appreciate Zhiliang Fan, Liang Tang, Dihong Tian, Zhiyuan Zhan, Guangfan Zhang, Qing Zhao and their family members for their friendship, which made my life in Atlanta much more enjoyable.

Last but not least, I would like to express my deepest gratitude to my family. I would like to thank my maternal grandmother and my parents for their unconditional and eternal love and support. Their encouragement is always with me and has given me the strength to complete this work. Special thanks to my wife, Ke Wang, for her enthusiasm in life and work and for her love and patience throughout the years. One of the best experiences that we lived through during this period was the birth of our son Marco Tianyu Chen, whose sunny smiles made the thesis writing a fun part of my Ph.D. journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- ACPR: Adjacent Channel Power Ratio
- BER: Bit Error Rate
- CCDF: Complementary Cumulative Distribution Function
- CDMA: Code Division Multiple Access
- CF: Crest Factor
- CSI: Channel State Information
- DFT: Discrete Fourier Transform
- FFT: Fast Fourier Transform
- FIR: Finite Impulse Response
- IAR: Instantaneous-to-Average Power Ratio
- IBI: Inter-block Interference
- ISI: Inter-symbol Interference
- ML: Maximum Likelihood
- MSE: Mean-square Error
- OFDM: Orthogonal Frequency Division Multiplexing
- PA: Power Amplifier
- PD: Predistorter

- PAR: Peak-to-Average Power Ratio
- PTAM: Pilot Tone Assisted Modulation
- PSD: Power Spectral Density
- PTS: Partial Transmit Sequence
- SLM: Selected Mapping
- SLMTC: Selected Mapping with Threshold and Clipping
- SNR: Signal-to-Noise Ratio
- SPTM: Selected Pilot Tone Modulation
- TIC: Transmitter Induced Cyclostationarity
- TV: Time-varying

SUMMARY

As wireless communication becomes an ever-more important and pervasive part of our everyday life, system capacity and quality of service issues are becoming more critical. In order to increase the system capacity and improve the quality of service, it is necessary that we pay closer attention to bandwidth and power efficiency issues.

Orthogonal Frequency Division Multiplexing (OFDM) is a multicarrier modulation technique for high speed data transmission over multipath fading channels and is generally regarded as bandwidth efficient. However, the price paid for the high bandwidth efficiency is low power efficiency. OFDM signals suffer from high peak-to-average power ratios (PARs) which lead to power inefficiency in the RF portion of the transmitter. Moreover, in OFDM, the well-known pilot tone assisted modulation (PTAM) technique utilizes a number of dedicated training pilots to acquire the channel state information (CSI), resulting in somewhat reduced bandwidth efficiency.

In this dissertation, we will address the above mentioned bandwidth efficiency and power efficiency issues in wireless transmissions. To avoid bandwidth efficiency loss due to dedicated training, superimposed training has been proposed in the literature to identify frequency selective fading channels. In superimposed training, the training pilots are "added onto" the information data instead of being "inserted into" the data stream as in conventional dedicated training; thus, superimposed pilots do not result in bandwidth reduction. However, since the pilots and the data are "co-mingled" in superimposed training, interference from the unknown data degrades the accuracy of the channel estimate, thus increasing the bit error rate (BER). In this dissertation, we will first develop a superimposed training framework that can be used to track the frequency selective as well as the Doppler shift characteristics of a channel. This scheme can be applied to any block transmission system. Later on, we will propose a generalized superimposed training framework that allows improved channel estimates with controllable interference from the data.

Another important modern transmission format is code division multiple access (CDMA), a popular platform on which 3G cellular networks are built. In forward link CDMA, a set of orthogonal codes are added together before transmission. Therefore, forward link CDMA signals have high PARs, similar to the situation with OFDM where data from multiple subcarriers are summed.

Power inefficiency becomes a critical issue with high PAR signals, such as OFDM and forward link CDMA. Many PAR reduction techniques have been proposed in the literature. Among them, selected mapping (SLM) is one of the most attractive approaches as it is distortionless and many variations of SLM can be formulated. However, two issues arise in SLM: high computational complexity and reduced bandwidth efficiency due to the need to share the side information with the receiver.

In this dissertation, we focus on the SLM framework and investigate its complexity and side information issues for both OFDM and forward link CDMA. We first propose a dynamic SLM algorithm to greatly reduce the computational requirement of SLM without sacrificing its PAR reducing capability. In stead of exhaustive search, adaptive search is carried out with dynamic SLM. The dynamic SLM idea can be applied to both OFDM and CDMA. However, to avoid the explicit transmission of side information, we need to deal with different signal formats differently. We propose a number of blind SLM techniques for OFDM and for forward link CDMA; they avoid the transmission of any side information and are easy to implement. Our proposed blind SLM technique for OFDM is a novel joint channel estimation and PAR reduction algorithm, for which bandwidth efficiency - power efficiency - complexity - bit error rate tradeoffs are carefully considered.

CHAPTER 1

INTRODUCTION

1.1 Motivations

Orthogonal frequency division multiplexing (OFDM) and code division multiple access (CDMA) have gained tremendous popularity in the past two decades for wireless data and voice transmissions.

OFDM is robust against inter-symbol interference (ISI) and permits a simple one-tap frequency-domain equalizer. Therefore, it has been adopted by many high-speed data transmission standards, such as digital audio broadcast (DAB) and digital video broadcast (DVB) in Europe, as well as in asymmetric digital subscriber line (ADSL), WLAN (IEEE 802.11a/g) and WiMAX (IEEE 802.16) applications. On the other hand, CDMA has provided greater capacity for voice and data communications, and is the common platform on which 3G cellular networks are built.

However, in wireless transmissions, reduced bandwidth efficiency usually results when one attempts to acquire the wireless channel state information (CSI) using conventional training methods. Multipath and Doppler effects are two major impairments that wireless communications signals encounter during their propagation. Obviously, one has to accurately estimate the CSI before carrying out channel equalization. Dedicated training is commonly used where a known training signal (called pilot) is transmitted before or interleaved with the unknown information data. However, pilots consume valuable bandwidth that otherwise can be used to transmit data. For example, in the FUSC (full usage of subchannels) mode of a forward link WiMAX system (IEEE 802.16e-2005 version), 42 out of 426 usable subcarriers are allocated to pilots and the rest to data, rendering an approximate 10% of bandwidth efficiency loss. On the other hand, superimposed training has been considered in [12, 13, 19, 32, 43, 51, 67, 77, 79] as an alternative to the conventional training techniques: pilots are added *onto* the data and the CSI is estimated without sacrificing the

data rate.

Superimposed training; i.e., simultaneous information transfer and channel sounding, is not a new idea. A 1965 reference [32] describes such a system, albeit for analog communication. In wireline communication, reserving dedicated slots for training makes a lot of sense, as channel estimation is done "once and for all." In a mobile communication environment however, training constitutes a recurring overhead, which should be avoided if possible. In contrast, superimposed training does not take up bandwidth. The superimposed training framework can be generalized to include precoding, and the so-called "affine precoding" structure has generated a lot of recent interest [41], [10], [50].

Another important consideration for the wireless transmission system is its power efficiency. OFDM and forward link CDMA signals suffer from high peak-to-average power ratios (PARs), making them susceptible to nonlinearities that are inherent in the RF/microwave power amplifiers (PAs). To avoid nonlinear distortions, either very linear PAs have to be used, or the input signal must be backed-off to well below the 1 dB compression point. Both options are power inefficient.

Power inefficiency leads to low battery life for the mobile user and high operating cost for the base station, and has become a critical issue for wireless communication applications. In a typical cellular base station, the RF PA and its associated cooling equipment are responsible for approximately 50% of the overall DC power consumption and 60% of its physical size [64]. With the convergence of computing and communications, a variety of multimedia features will be offered to modern wireless handsets. As compared with traditional voice services, a higher demand is placed on the battery. Advancing battery technologies alone will not meet the growing demand for power in consumer electronics, since unfortunately, Moore's Law does not seem to apply to batteries [37]. It is reported that, in today's cellular phones, over 90% of the power used to transmit the signal is wasted in the form of heat that stays inside the phone [1]. Thus, research on PAR reduction is very well motivated.

1.2 Objective

The objective of this dissertation is to provide a suite of relatively simple but effective solutions to solve the bandwidth efficiency and power efficiency problems in modern wireless transmission systems. More specifically, this dissertation will focus on the following two topics:

- Channel estimation using superimposed training
- PAR reduction using selected mapping

Most of the literature on superimposed training focuses on the estimation a frequency selective channel, which is commonly modeled as a finite impulse response (FIR) filter. However, in some wireless scenarios, e.g., in mobile applications, a frequency selective time varying (doubly selective) channel model is more realistic due to the presence of Doppler shifts. Building on a basis expansion model for the doubly selective channel, we will address the estimation of multipath weighting coefficients as well as Doppler frequencies of the doubly selective channel model under the superimposed training framework.

One issue with superimposed training is the estimation performance degradation caused by the unknown data acting as noise. We will also investigate techniques that can suppress data induced distortion in superimposed training.

There has been a great deal of research on PAR reduction for OFDM signals. Among the various PAR reduction techniques, selected mapping (SLM), first proposed by Bäuml, Fischer and Huber in 1996 [6], is one of the most attractive approaches. SLM is a distortionless and effective PAR reduction method. However, two issues arise in SLM: high computational complexity and reduced bandwidth efficiency due to the transmission of side information. In this dissertation, we will try to address the above mentioned issues for SLM in both OFDM and CDMA applications.

1.3 Outline

The rest of this dissertation is organized as follows:

In Chapter 2, we will review the literature on channel estimation and on PAR reduction. For channel estimation, we utilize a basis expansion model for doubly selective channels and demonstrate the advantages of the superimposed training framework. For PAR reduction, we first analyze the roles that PA linearization and PAR reduction play in improving the power efficiency of a nonlinear PA; we then describe the conventional SLM algorithm for OFDM and point out its certain deficiencies.

In Chapter 3, we will focus on the estimation and equalization of the basis expansion time-varying channel model using superimposed training. We propose a novel technique that estimates not only the channel tap coefficients but also the Doppler frequencies for the block transmission system.

In Chapter 4, we will first propose a dynamic SLM scheme to greatly reduce the computational complexity of the conventional SLM without sacrificing the PAR reducing capability. We then propose three blind SLM techniques, two for forward link CDMA and one for OFDM; these methods avoid the transmission of any side information and add little complexity to the transmitter or the receiver.

In Chapter 5, we will further explore the blind SLM technique for OFDM that is proposed in Chapter 4, from the view point of joint channel estimation and PAR reduction. We will examine the bandwidth efficiency - power efficiency - complexity - bit error rate tradeoffs that are involved.

Finally, in Chapter 6, we will summarize this dissertation and suggest topics for future research.

For the reader's convenience, we have attempted to keep every chapter as self contained as possible.

CHAPTER 2

BACKGROUND

2.1 Bandwidth Efficiency and Channel Estimation

In this section, we first introduce a simple but effective basis expansion channel model, and then review existing channel estimation techniques for this model. We then discuss the bandwidth efficiency issue with conventional dedicated training methods, and present novel results by resorting to superimposed training.

2.1.1 Basis Expansion Channel Model

Wireless channels can be characterized by "multipath reception" and/or "Doppler spread". The specific channel characteristics can be affected by many factors, such as materials of the surrounding objects, weather and relative velocity etc. For example, in an indoor radio environment, there are usually multiple reflected signals which cause the ISI and the line-of-sight signal path may or may not be present. Sometimes, not only the multipath effect but the Doppler spread must also be taken into account, the latter having an impact on synchronization of the receiver. A frequency and time selective (doubly selective) channel model can accommodate both effects.

Assuming that the channel is linear, a generic discrete-time equivalent baseband input-output data model can be written as (see e.g., [65], [24], [40]),

$$y[n] = \sum_{l=0}^{L-1} h[n;l]x[n-l] + v[n], \tag{2.1}$$

where $\{x[n]\}_{n=0}^{N-1}$ is the transmitted signal, $\{y[n]\}_{n=0}^{N+L-1}$ is the received signal, $h[n;l]$ is the time-varying (TV) impulse response that is the convolution of the transmit filter, the TV channel, and the receive filter, and $v[n]$ is zero-mean additive noise. When $h[n;l] = h[l]$, $\forall n$, the channel is frequency selective but time non-selective. If $h[n;l] = 0$, $\forall l \neq 0$, the channel is time selective but frequency non-selective. Since in (2.1), $h[n;l]$ is allowed to vary in both

n and l, we encounter a more general frequency and time (and thus doubly) selective fading environment.

As pointed out in [65], if $h[n;l]$ changes slowly with n, one may resort to adaptive algorithms to track the channel variations. However, if the channel is fast changing, modeling the time variations in $h[n;l]$ is necessary so the problem is not ill-posed. In [65], [24], under the assumption that the multipath effect is caused by a small number of distinct reflectors, and the motion of the receiver is constant, a complex exponential basis expansion model can be formulated for the TV channel; i.e.,

$$h[n;l] = \sum_{q=1}^{Q} h_q[l]\ e^{j\omega_q n}, \tag{2.2}$$

where $j = \sqrt{-1}$, ω_q is proportional to the carrier frequency, path velocity, and the sampling period [24].

Another commonly used channel model is the frequency selective block fading model for block transmissions such as OFDM. Block fading means $h[n;l]$ can be considered time-invariant over the i-th block; i.e., $h[n;l] = h_i[l]$ for $(i-1)N \leq n \leq iN-1$. For the rest of the dissertation, we will drop the block index i for notational simplicity, since most block transmission systems can be made free of inter-block interference (IBI) by employing the cyclic prefix or zero padding. Thus, the frequency selective block fading channel model can be represented as $h[l]$ for $0 \leq l \leq L-1$.

2.1.2 Superimposed Training

In [65], a training based approach is used to estimate $\{h_q[l]\}$, and the cycles $\{\omega_q\}$ are detected by a 4th-order statistic. In [24], blind estimation of $h[n;l]$ is carried out relying on channel diversity; i.e., a single input multiple output (SIMO) scenario is assumed. Recently in [67] and [77], the superimposed training technique of [20], [42], [80] for frequency selective channels is extended to the doubly selective model (2.2). With superimposed training, it is possible to estimate the channel without slowing the information rate.

Superimposed training can be regarded as a transmitter-induced cyclostationarity (TIC) method. In TIC, cyclostationarity can be enforced at the transmitter by means of coding

(repetition codes) [66]; filtering (analysis-synthesis filterbanks) [23]; modulation with deterministic periodic sequences [15, 57]; or superposition of periodic pilot sequences [19, 42, 80]. TIC-based approaches are appealing for three reasons: (i) they do not require dedicated slots for training and thus achieve higher bandwidth efficiency compared with traditional training-based approaches; (ii) they entail lower computational complexity and smaller variance by employing second- (or even first- [80]) as opposed to higher-order statistics of the data; (iii) they acquire the CSI with no constrains on the channel zero locations as would be the case with receiver-induced cyclostationarity methods. Among the TIC-based approaches, superimposed training has attracted a lot of attention lately due to its simplicity and better CSI estimation accuracy.

The idea of superimposed training; i.e., simultaneous information transfer and channel sounding, was first described in a 1965 paper [32], albeit for analog communications. It was advocated for digital communication systems by Farhang-Boroujeny in 1995 [19] and more investigations followed in [12, 13, 43, 51, 67, 77, 79]. Recently, the superimposed training framework has been generalized to include precoding (affine precoding) and has attracted much interest [41, 50]. The design of optimal superimposed training sequences has also been considered in [22, 51] by minimizing the mean-squared error (MSE) of the channel estimate.

2.1.3 Bandwidth Efficiency

OFDM is gaining much popularity due to its high spectral (bandwidth) efficiency and the simple equalizer structure to combat the ISI. However, to enable frequency-domain equalization, one has to estimate the CSI first.

Pilot tone assisted modulation (PTAM) is commonly employed to acquire the CSI in OFDM whereby training pilots or pilot tones are inserted in the time or frequency grids of the OFDM symbols. However, these pilots consume valuable bandwidth and reduce the data rate. For instance, consider the case with $N = 64$ subcarriers, a frequency selective block fading channel of $L = 8$ taps, and a cyclic prefix of length $G = 7$ (so that $G \geq L - 1$). The minimum number of pilot tones is $L = 8$, giving rise to a bandwidth efficiency of $\frac{N-L}{N+G} = 78.9\%$. On the other hand, in superimposed training, the pilots are added *onto* the

data and the CSI is estimated without sacrificing the data rate (e.g., bandwidth efficiency of $\frac{N}{N+G} = 90.1\%$ for the above example).

In Chapter 3, we will investigate the superimposed training method for doubly selective channels, and propose a novel technique for estimating the frequencies $\{\omega_q\}$. In [67], $\{\omega_q\}$ are assumed known. With known $\{\omega_q\}$, the time-invariant portion of the channel, $\{h_q(l)\}$, is recovered via linear least squares in [67]. We will present an alternative way of estimating $\{h_q(l)\}$ by examining the Fourier Transform (FT) of appropriately constructed sub-processes. We will also propose two equalizers to recover the input symbols. The superimposed training technique is superior to the explicit training approach of [65] in that no dedicated slots for training are required. It is also advantageous over the blind approach of [24] since no receiver diversity is sought to ensure channel identifiability.

2.2 Power Efficiency and PAR Reduction

In this section, we first explain how PAR reduction can improve the power efficiency of a wireless communication system, and then review various PAR reduction techniques. We offer a brief introduction to SLM, and describe existing techniques to reduce the computational complexity of SLM as well as schemes that try to avoid the transmission of the side information.

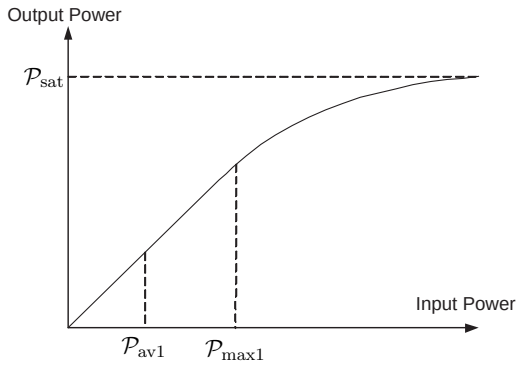
Consider the input-output characteristic of an ideal linear (or linearized) PA shown in Fig. 1(a). We adopt the following notations:

$x(t)$,	the baseband PA input;
$\mathcal{P}_{\text{sat}}$,	the maximum output power that the PA is capable of producing;
$\mathcal{P}_{\max} = \max_{0\le t\le T} \lvert x(t)\rvert^2$,	the maximum input power;
$\mathcal{P}_{\text{av}} = \bar{E}[\lvert x(t)\rvert^2]$,	the average input power;

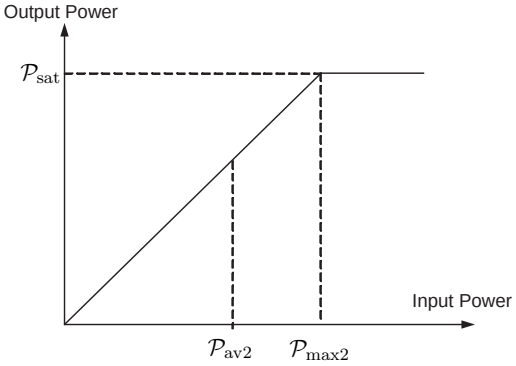
where T is the block duration for block transmission and $\bar{E}[\cdot]$ denotes expectation, or time-averaged expectation if $x(t)$ is nonstationary.

The peak-to-average power ratio (PAR) is a characteristic of the input signal and is defined as [47, 62]

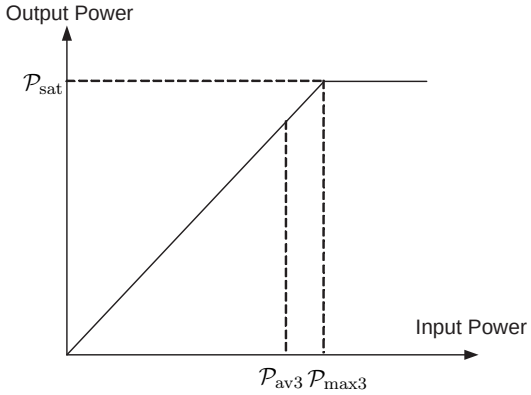
$$\text{PAR}(x(t)) = \frac{\mathcal{P}_{\max}}{\mathcal{P}_{\text{av}}} = \frac{\max_{0\le t\le T} \lvert x(t)\rvert^2}{\bar{E}[\lvert x(t)\rvert^2]}, \tag{2.3}$$



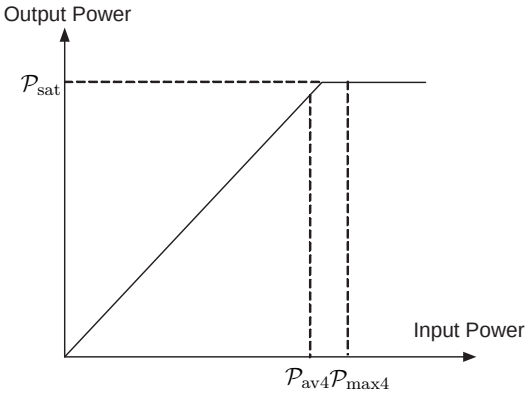
(a) Nonlinear PA with input back-off.
$\text{PAR}_1(\text{dB}) = \mathcal{P}_{\text{max1}}(\text{dB}) - \mathcal{P}_{\text{av1}}(\text{dB})$.



(b) Ideal linear PA.
$\text{PAR}_2(\text{dB}) = \mathcal{P}_{\text{max2}}(\text{dB}) - \mathcal{P}_{\text{av2}}(\text{dB})$.
$\text{PAR}_2 = \text{PAR}_1$.
$\mathcal{P}_{\text{max2}} > \mathcal{P}_{\text{max1}}$, $\mathcal{P}_{av2} > \mathcal{P}_{av1}$.



(c) After PAR reduction.
$\text{PAR}_3(\text{dB}) = \mathcal{P}_{\text{max3}}(\text{dB}) - \mathcal{P}_{\text{av3}}(\text{dB})$.
$\text{PAR}_3 < \text{PAR}_2$.
$\mathcal{P}_{\text{max3}} = \mathcal{P}_{\text{max2}}$, $\mathcal{P}_{\text{av3}} > \mathcal{P}_{\text{av2}}$.



(d) Allow occasional saturation.
$\text{PAR}_4(\text{dB}) = \mathcal{P}_{\text{max4}}(\text{dB}) - \mathcal{P}_{\text{av4}}(\text{dB})$.
$\text{PAR}_4 = \text{PAR}_3$.
$\mathcal{P}_{\text{max4}} > \mathcal{P}_{\text{max3}}$, $\mathcal{P}_{\text{av4}} > \mathcal{P}_{\text{av3}}$.

Figure 2.1: PA linearization and PAR reduction can improve the PA efficiency by reducing the amount of back-off that is needed.

or in dB ($10\log 10$) scale

$$\text{PAR(dB)} = \mathcal{P}_{\max}(\text{dB}) - \mathcal{P}_{\text{av}}(\text{dB}). \tag{2.4}$$

The crest factor (CF) is the square root of the PAR so they have the same value in dB and are often used interchangeably in the literature.

In Fig. 1(a), the PA is linear up to $\mathcal{P}_{\max 1}$, but is nonlinear afterwards. Nonlinearity generates in-band distortion as well as adjacent channel interference. To avoid those detrimental nonlinear effects, the input signal is often backed-off to the PA's linear region as shown in Fig. 1(a). The corresponding power efficiency is very low, often in the range of 10% or much less [16].

With PA linearization, we strive to achieve an *ideal* linear input-output characteristic shown in Fig. 1(b). The input signal is amplified undistorted until $\mathcal{P}_{\text{sat}}$ is reached. In Fig. 1(b), the average input power is higher than that in Fig. 1(a); i.e., $\mathcal{P}_{\text{av}2} > \mathcal{P}_{\text{av}1}$, demonstrating how power efficiency can be improved via PA linearization.

Among the various PA linearization architectures, adaptive baseband predistortion is among the most cost-effective. We are interested in the memory polynomial predistorter (PD) model given by [18]

$$x_{PD}[n] = \sum_{k=1}^{K} \sum_{q=0}^{Q} a_{kq} x[n-q] \big| x[n-q] \big|^{k-1}, \tag{2.5}$$

where $x[n] = x(t)|_{t=n/F_s}$ is the sampled version of the input $x(t)$ with sampling frequency F_s, $x_{PD}[n]$ is the discrete-time output of the predistorter, and $\{a_{kq}\}$ are the predistorter coefficients. This predistorter has memory depth Q and highest nonlinearity order K. The indirect learning architecture is used to solve for the parameters $\{a_{kq}\}$ via linear least squares; see [18] for details. Note that when $Q = 0$, eq. (2.5) becomes a memoryless polynomial predistorter, which may be sufficient for memoryless PAs, such as handset PAs with narrowband inputs.

If we can reduce the PAR of the input signal from PAR_2 to PAR_1, we arrive at a situation depicted in Fig. 1(b). The peak power is the same as in Fig. 1(b), but thanks to PAR reduction, the average input power is increased; i.e., $\mathcal{P}_{\text{av}2} > \mathcal{P}_{\text{av}1}$, boosting the efficiency of the PA.

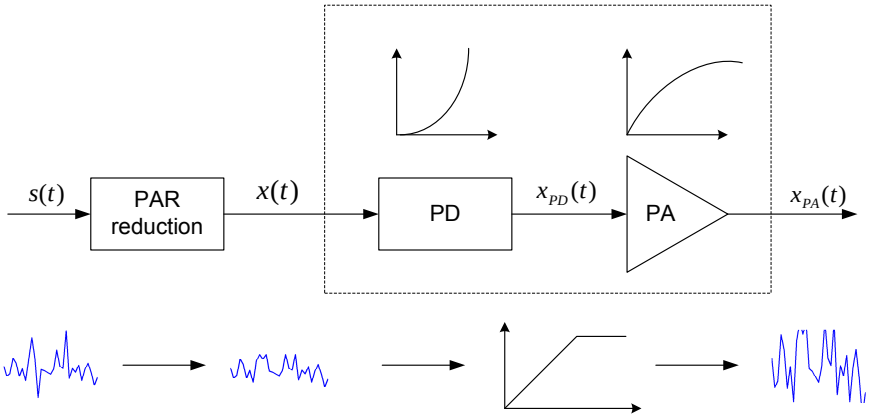


Figure 2.2: Power efficiency enhancement by PAR reduction and PA linearization.

If we drive the PA harder by scaling up the input so the signal occasionally enters the saturation region of the PA, we can achieve even higher efficiency at the expense of nonlinear distortions.

Joint PAR reduction and PA linearization have been investigated in [55] and [60]. In [55], the authors investigated the BER performance degradation due to inaccuracy of the side information used by the PAR reduction for a multicarrier CDMA system, but gave no details of PA linearization. In [60], a commercial chip that implements deliberate clipping was used as the PAR reduction preprocessor and a lookup table was used for PA linearization.

In Chapter 4, we demonstrate by testbed experiments that the combination of PAR reduction and PA linearization can significantly improve the transmission power efficiency. PA linearization usually functions regardless the input signal format (e.g., OFDM vs. CDMA), but many PAR reduction algorithms are developed with a particular type of signal in mind. The overall PAR reduction – predistortion linearization schematic is shown in Fig. 2.2.

2.2.1 PAR Reduction Techniques for OFDM and CDMA

Many PAR reduction techniques have been proposed in the literature. Most of them deal with OFDM signals; the relevant PAR reduction techniques include (i) distortionless PAR reduction, such as coding [17, 30, 61], tone reservation [62], tone injection [62], active constellation extension [34], selected mapping [6, 44] and partial transmit sequence [45]; (ii) PAR reduction with distortion, such as deliberate clipping [31, 38], transmit filtering [59],

companding [29, 71] approaches etc; and (iii) various combinations of the above. These methods entail different PAR reducing capability - complexity - information rate - bit error rate (BER) tradeoffs.

Deliberate clipping and companding are the simplest PAR reduction algorithms [38, 71]. Clipping reduces the PAR by limiting the peak power to a predetermined threshold. In companding, the dynamic range of the input signal is reduced by way of a known compressing function. These methods, however, cause increase in the BER and/or spectral regrowth. In ref. [30], codewords with low PARs were selected. However, the approach of reducing PAR through coding requires extensive search for good codes and is limited to multicarrier systems with a small number of subcarriers. Tone reservation and tone injection [62] are considered distortionless PAR reduction methods, but the continuous parameter optimization involved can be computationally very intensive. Among the PAR reduction algorithms, SLM and partial transmit sequence (PTS) are attractive [6, 44, 45] due to their substantial PAR reducing capability without sacrifice in signal fidelity. SLM selects one signal with the lowest PAR from a set of "equivalent" signal representations that are related in the frequency domain by a series of phase rotations. PTS divides the data into sub-blocks and optimizes the phase for each sub-block so that the combined signal has a low PAR. Besides high computational complexity, SLM and PTS also require the transmission of the side information regarding the optimum phase selection; this side information transmission issue may hinder their practical use in high speed wireless applications. In this dissertation, we will focus on the SLM framework and propose novel techniques to avoid the transmission of side information.

In comparison with OFDM, the body of literature on PAR reduction for CDMA signals is rather small. In [35, 58], a Walsh code selection algorithm was proposed to reduce the PAR by assuming that only part of the channels are active at any given time. Based on the same assumption, a PAR reduction algorithm was proposed in [68] by adding a signal that is orthogonal to all the active channel codes. In [36], the authors proposed to reduce the PAR of the forward link CDMA signal by changing the signs of the Walsh codes in one branch of the quadrature modulation. In [49, 76], SLM and PTS were adopted for PAR

reduction of OFDM-CDMA systems; however, both techniques were applied only to the OFDM part and not to the CDMA part of the system.

Although PAR reduction for OFDM and that for CDMA can share some common grounds, the techniques themselves are largely different because of the different signal formats. Moreover, the PAR definitions are also different. OFDM is a block processing technique; therefore, the peak is well defined over a block (c.f. 2.3). In OFDM, PAR is treated as a random variable; its complementary cumulative distribution function (CCDF) is often analyzed to assess the performance of a PAR reduction algorithm. In CDMA however, the instantaneous-to-average power ratio (IAR) is of interest [36]. Its CCDF is computed and the PAR is taken as the value γ such that the CCDF $\Pr(\text{IAR} > \gamma) = p$ for a small p [5]. Therefore, we propose different techniques for OFDM and CDMA in Chapter 4 though all these proposed methods are based on the SLM concept.

Next, we review SLM and related PAR reduction techniques for OFDM.

2.2.2 Selected Mapping for OFDM

Denote by $\{X[k]\}_{k=0}^{N-1}$ one block of the frequency domain OFDM signal drawn from a known constellation, where N is the number of sub-carriers. Note that the block index is dropped here since OFDM can be free of IBI with proper use of the cyclic prefix. The complex baseband time-domain OFDM signal can be written as

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k]\ e^{j2\pi kt/T_s}, \quad 0 \le t \le T_s, \tag{2.6}$$

where T_s is the OFDM symbol period and $j = \sqrt{-1}$. The PAR of $x(t)$ can be defined as in (2.3) with $T = T_s$.

The Nyquist-rate sampled time-domain OFDM signal is $x[n] = x(t)|_{t=nT_s/N}$, whose PAR has been shown in [62] to have the following CCDF:

$$\Pr\{\text{PAR}(x[n]) > \gamma\} = 1 - (1 - e^{-\gamma})^N. \tag{2.7}$$

Since the input to the PA must be a continuous-time signal, the CCDF of the PAR of $x(t)$ is of interest. Oversampling has been used to approximate the CCDF of the PAR of the continuous-time OFDM signal by introducing an empirical modification [69]. In [48,74],

Rice's level-crossing rate analysis [54] has been applied to the study of the continuous-time PAR distribution. In [72], extreme value theory is employed to provide a simple but accurate approximation of the CCDF of PAR($x(t)$):

$$\Pr\{\text{PAR}(x(t)) > \gamma\} = 1 - \exp\left\{-e^{-\gamma} N \sqrt{\frac{\pi}{3}\ln N}\right\}. \tag{2.8}$$

In SLM, besides $x(t)$, a set of "equivalent" representations of $x(t)$ are first generated, and the one with the lowest PAR is selected for transmission. Denote by $\phi^{(m)}[k],\ 0 \le k \le N-1,\ 0 \le m \le M-1$, a set of M (random) phase sequences. Table 2.1 illustrates how one such phase table might look; it is a known table that is available to both the transmitter and the receiver. Each row (indexed by m) represents a different phase rotation sequence; the column index k corresponds to the frequency-domain subcarriers.

Table 2.1: Example of an $M \times N$ phase rotation table that is available to both the transmitter and the receiver.

	$k=0$	$k=1$	$k=2$	$k=3$	$\dots$	$k=N-1$
0	2.8282	2.4586	0.7253	1.4967		2.7361
1	−1.6893	1.6468	−0.3474	2.6503		2.6195
⋮	⋮		⋮		⋮	
m	0.6713	−0.2735	0.7253	−2.0341		−0.5638
⋮	⋮		⋮		⋮	
$M-1$	−0.0881	−3.0253	1.8343	−0.5925		2.4734

In SLM, we first rotate the phases of $X[k]$ as described by

$$Z^{(m)}[k] = X[k]e^{j\phi^{(m)}[k]}. \tag{2.9}$$

It is clear that $Z^{(m)}[k]$ and $X[k]$ contain the same information, but their time-domain counterparts $z^{(m)}(t)$ and $x(t)$ can have very different PAR values. From the M candidate $z^{(m)}(t)$ signals, $z^{(\bar{m})}(t)$, which has the lowest PAR, is transmitted; i.e.,

$$\text{PAR}\left(z^{(\bar{m})}(t)\right) = \min_{1\le m\le M} \text{PAR}\left(z^{(m)}(t)\right). \tag{2.10}$$

Given Table 2.1 and $\bar{m}$, the receiver can recover $X[k]$ by reversing the operation in (2.9). SLM is a distortionless PAR reduction method.

Table 2.2: SLM algorithm.

Step 0	Set $\bar{m} = m = 1$.
Step 1	Form $z^{(m)}(t)$, and compute $\text{PAR}(z^{(m)}(t))$.
Step 2	If $\text{PAR}(z^{(m)}(t)) < \text{PAR}(z^{(\bar{m})}(t))$, then go to Step 2a; else go to Step 2b.
	Step 2a Set $\bar{m} = m$.
	Step 2b $m = m + 1$.
Step 3	If $m > M$, then go to Step 4; else go to Step 1.
Step 4	Transmit $z^{(\bar{m})}(t)$.

Optimal design of the phase table for SLM has been investigated in [78]. If $\{\phi^{(m)}[k]\}$ are i.i.d. satisfying $E[e^{j\phi^{(m)}[k]}] = 0$, then the best SLM performance can be achieved; the corresponding CCDF is given by

$$\Pr\{\text{PAR}(z^{(\bar{m})}(t)) > \gamma\} = [\Pr\{\text{PAR}(x(t)) > \gamma\}]^M . \tag{2.11}$$

It is pointed out in [78] that the simplest and yet optimal phase rotation table is one that consists of 0 and π entries with equal probability. In that case, no multiplication is necessary in (2.9) since $e^{j0} = 1$ and $e^{j\pi} = -1$. The phase rotation table is pre-determined and is stored at both the transmitter and the receiver, so real-time optimization of the phase sequence is not necessary.

The step-by-step SLM algorithm is summarized in Table 2.2.

SLM is one of the most attractive PAR reduction techniques due to its distortionless nature and its great PAR reducing capability as revealed by (2.11). However, two drawbacks of SLM need to be addressed: high computational complexity due to the $M - 1$ additional IFFTs and reduced bandwidth efficiency due to the transmission of side information $\bar{m}$.

2.2.2.1 Computational Complexity of SLM

From Table 2.2, M mappings per OFDM block are required, consuming approximately M times the computational resources as compared to the original OFDM system.

To reduce the computational requirement of SLM, a simple approximation of the inverse discrete Fourier transform (IDFT) is proposed in [70], but the price paid is degradation in the PAR reducing capability.

In Section 4.1, we introduce a dynamic SLM technique, which greatly reduces the computational requirement of SLM without sacrificing its PAR reducing capability.

2.2.2.2 Bandwidth Efficiency of SLM

In SLM, the optimum phase sequence index $\bar{m}$ ($\log_2 M$ bits) needs to be transmitted as side information, which is of critical importance to the receiver for decoding and is generally protected by channel coding [6]. To avoid the information rate loss caused by the transmission of $\bar{m}$, several blind SLM schemes have been proposed. In [8], a scrambling technique was described. A $\log_2 M$-bit binary label is inserted as prefix to the frequency-domain OFDM signal and passed through a scrambler. Since the selected label is used in the receiver implicitly during descrambling, an erroneous reception of the label bits does not affect the error performance. However, additional complexity is introduced to both the transmitter and the receiver due to the scrambling process, and the $\log_2 M$-bit binary label is still transmitted as side information. In [28], a blind SLM receiver was proposed by employing a maximum likelyhood (ML) decoder, which avoids the transmission of any side information. Unfortunately, the ML decoder imposes a high computational demand at the receiver and the overall BER performance is degraded due to the occasionally erroneous detection of $\bar{m}$.

In Section 4.2, we propose to embed the SLM index information in the locations of the pilot tones that are used for channel estimation. At the receiver, the SLM index can be simply detected by a peak power detector after the FFT operation. The robustness of the proposed blind SLM algorithm will be demonstrated through simulation comparisons against the method in [28].

2.3 Joint Bandwidth Efficiency and Power Efficiency Enhancement

The blind SLM algorithm to be proposed in Section 4.2 is essentially a joint channel estimation and PAR reduction scheme; we refer to it as blind selected pilot tone modulation (BSPTM), which combines SLM and PTAM. However, PTAM requires dedicated pilot subcarriers for channel estimation. To further improve the bandwidth efficiency of BSPTM, we replace the PTAM part in Chapter 5 with superimposed training. The selection of the optimal pilot tones will also be discussed in Chapter 5.

The concept of joint channel estimation and PAR reduction has also been studied in [21, 26]. However, both approaches focus on the design of the pilot tones and have limited PAR reducing capabilities.

CHAPTER 3

IMPROVING BANDWIDTH EFFICIENCY BY SUPERIMPOSED TRAINING

In this chapter, we adopt the basis expansion model (c.f., eq. (2.2)) for doubly selective fading channels and adopt the superimposed training framework to acquire the channel state information. Specifically, we add a known periodic sequence onto, instead of multiplexing it into, the information sequence. There is no reduction in transmission rate, and it is possible to estimate the time-varying channel without requiring receiver diversity. We propose a novel Doppler frequency estimator and describe two equalizers for symbol recovery. Bit error rate performance is demonstrated for the iterative joint data - channel estimator. Tradeoffs regarding power allocation to the pilots and to the data are also explored. We will not limit ourselves to OFDM systems as our proposed approach can be applied to any block transmission systems with a guard interval equal to or longer than the equivalent maximum channel order $L-1$.

Differences between our proposed approach and the one described in [67] are: (i) we do not assume $\{\omega_q\}$ to be known; (ii) we adopt a single input single output (SISO) (instead of SIMO) formulation; (iii) we show equalization performance whereas [67] focuses on the channel identification problem; (iv) we explore experimentally, the issue of optimum power allocation among the pilots and the symbols.

3.1 Channel Estimation

Denote by $s[n]$ the zero-mean information sequence. Instead of transmitting $x[n] = s[n]$, we transmit $x[n] = s[n]+b[n]$ under the superimposed training framework [20], [42], [80], where $b[n]$ is a known sequence (also called a pilot) that is added onto, instead of multiplexed into, $s[n]$.

3.1.1 Frequency Selective Block Fading Channel Estimation

If the channel is time-invariant within a block of N samples, eq. (2.1) becomes

$$y[n] = (s[n] + b[n]) * h[n] + v[n] = d[n] + u[n], \tag{3.1}$$

where $*$ denotes linear convolution,

$$d[n] = b[n] * h[n], \quad \text{and} \quad u[n] = s[n] * h[n] + v[n].$$

Therefore, the "output-only" system identification problem of finding $h[n]$ from $y[n]$ in the presence of additive noise $v[n]$, can be recast into an "input-output" system identification problem of recovering $h[n]$ based on the "input" $b[n]$ and output $y[n]$ in the presence of additive "noise" $u[n]$. In fact, $b[n]$ can be arbitrary (periodic or aperiodic). Since only periodic $b[n]$ is considered in [20], [42], [80], we describe here a method of estimating

$$\mathbf{h} = [h[0],\ h[1], \ldots, h[L-1]]^T,$$

for a general (aperiodic) $b[n]$, where $(\cdot)^T$ denotes transpose.

Step 1. Solve for $\mathbf{h}$ using linear least squares:

$$\mathbf{h}_{LS} = (\mathbf{P}'\mathbf{P})^{-1}\ \mathbf{P}'\mathbf{y}, \tag{3.2}$$

where $'$ denotes Hermitian transpose, $\mathbf{P}$ is $(N+L-1)\times L$ Toeplitz with $[b[0],\ \mathbf{0}_{1\times(L-1)}]$ as its first row and $[b[0],\ \ldots,\ b[N-1],\ \mathbf{0}_{1\times(L-1)}]^T$ as its first column, and $\mathbf{y} = [y[0],\ y[1],\ \ldots,\ y[N+L-2]]^T$.

Step 2. Estimate the covariance matrix for $\mathbf{u} = [u[0],\ u[1], \ldots, u[N+L-2]]^T$ as

$$\mathbf{C} = \sigma_s^2\ \tilde{\mathbf{H}}\tilde{\mathbf{H}}' + \sigma_v^2\ \mathbf{I}, \tag{3.3}$$

where $\tilde{\mathbf{H}}$ is $(N+L-1)\times N$ Toeplitz with $[h_{LS}[0],\ \mathbf{0}_{1\times(N-1)}]$ as its first row and $[\mathbf{h}_{LS}^T,\ \mathbf{0}_{1\times(N-1)}]^T$ as its first column. A weighted least squares solution

$$\mathbf{h}_{WLS} = (\mathbf{P}'\mathbf{W}\mathbf{P})^{-1}\ \mathbf{P}'\mathbf{W}\mathbf{y}, \tag{3.4}$$

where the weight matrix $\mathbf{W} = \mathbf{C}^{-1}$, is often an improvement over $\mathbf{h}_{LS}$.

Step 3 (optional). Revise the $\mathbf{C}$ in (3.3) by replacing $\mathbf{h}_{LS}$ with $\mathbf{h}_{WLS}$, and solving (3.4) again. This step can be repeated, but often there is diminishing return in doing so.

An interesting problem to investigate is, for a fixed average power for the pilot,

$$\sigma_p^2 = \frac{1}{N}\sum_{n=0}^{N-1} |b[n]|^2, \tag{3.5}$$

what kind of $b[n]$ will result in the best BER performance without substantially increasing the peak-to-average power ratio of $x[n]$ (as compared to that of $s[n]$)? One drawback of an aperiodic $b[n]$ is that a large matrix inverse $\mathbf{W} = \mathbf{C}^{-1}$ is involved in (3.4). In [80], a simple periodic pilot sequence $b[n] = a\sum_k \delta[n - kP]$ is considered, where the period $P \geq L$ and $\delta(\cdot)$ is the Kronecker delta function. With such a $b[n]$, the equivalent of $\mathbf{h}_{LS}$ is obtained from simple "synchronized averaging", and no matrix inverse is needed.

3.1.2 Doubly Selective Channel Estimation

Let us now turn to the basis expansion model (2.2) for doubly selective channels. In order to ensure that the received data $y[n]$ is cyclostationary so the frequencies $\{\omega_q\}$ can be estimated in a straightforward manner, we consider $b[n]$ periodic with period $P \geq L$. For simplicity, let us use $P = L$ and assume that $R = N/L$ is an integer. Let us define $b_0[n]$ to be the first period of $b[n]$. If we write $n = rL + \tau$ with $0 \leq r \leq R-1$, $0 \leq \tau \leq L-1$, then $y_\tau[r] = y[rL + \tau]$ represents the rth length-L sub-record of $y[n]$. Substituting $n = rL + \tau$ and eq. (2.2) into $d[n] = \sum_{l=0}^{L-1} h[n; l]b[n - l]$, we obtain

$$d_\tau[r] = d[rL + \tau] = \sum_{q=1}^{Q} f_q[\tau]\ e^{jL\omega_q r}, \tag{3.6}$$

where

$$f_q[\tau] = e^{j\omega_q \tau}[h_q[\tau] \circledast b_0[\tau]], \tag{3.7}$$

and $\circledast$ denotes circular convolution.

Therefore, for a fixed τ, $d_\tau[r]$ of (3.6) consists of Q constant amplitude complex exponentials in r, with complex amplitudes $f_q[\tau]$ and frequencies $L\omega_q$.

Regarding

$$u[n] = \sum_{l=0}^{L-1} h[n; l]s[n - l] + v[n],$$

since both $s[n]$ and $v[n]$ are zero-mean, $u[n]$ is zero-mean as well. This implies that for a given τ, if we take the normalized DTFT of $y_\tau[r]$ w.r.t. r; i.e.,

$$Y_\tau[\alpha] = \frac{1}{R}\sum_{r=0}^{R-1} y_\tau[r]\ e^{-j\alpha r}, \qquad -\pi \le \alpha < \pi, \tag{3.8}$$

we expect to see peaks in $|Y_\tau[\alpha]|$ around $\alpha = L\omega_q$, for all τ with $f_q[\tau] \neq 0$. Here we assume that $|L\omega_q| < \pi$ to avoid any phase wrap-around.

If $|f_q[\tau]|$ is small (or zero) for a particular τ, the peak at $L\omega_q$ will not be seen in $Y_\tau[\alpha]$ for that τ. However, if we multiply (3.8) for all τ; i.e., we calculate

$$\mathcal{Y}[\alpha] = \prod_\tau Y_\tau[\alpha],$$

the peaks at $\{L\omega_q\}$ will be visible in $|\mathcal{Y}[\alpha]|$. Through multiplication, $|\mathcal{Y}[\alpha]|$ has a larger dynamic range than that for each $|Y_\tau[\alpha]|$ – the peaks at $\{L\omega_q\}$ are sharpened, and the noise floor is suppressed. The frequency estimates $\{L\hat{\omega}_q\}$ are easily obtained from the peak locations in $|\mathcal{Y}[\alpha]|$. As in all FT based harmonic retrieval problems, the variance of the $\hat{\omega}_q$ estimate can be shown to be on the order $O(N^{-3})$. Even if $f_q[\tau] = 0$, $Y_\tau(L\omega_q) \neq 0$ due to the finite data record and the noise present. Therefore, generally, one does not need to be concerned about the peak at $L\omega_q$ disappearing in $|\mathcal{Y}[\alpha]|$ due to a small $|f_q[\tau]|$. The example in Section 3.3.1 clearly illustrates these points. If the ω_q's are closely spaced (relative to $1/N$), high resolution techniques such as MUSIC (or root MUSIC) [56] can be considered.

Once the frequencies are estimated, we form

$$\hat{f}_q[\tau] = Y_\tau(L\hat{\omega}_q). \tag{3.9}$$

Based on (3.7), we can write

$$H_q(\omega) = F_q(\omega - \omega_q)/B_0(\omega),$$

where $H_q(\omega)$, $F_q(\omega)$, and $B_0(\omega)$ are the FT of $h_q[\tau]$, $f_q[\tau]$, and $b_0[\tau]$, respectively. Combining this with (3.9), it is then straightforward to estimate the time-invariant portion of the channel, $h_q[l]$.

If $b[n] = a\sum_k\ \delta(n - kL)$; i.e., $b[n]$ is a periodic delta function and thus $b_0[n] = a\ \delta[n]$, we find from (3.7),

$$\hat{h}_q[\tau] = [\hat{f}_q[\tau]/a]\ e^{-j\hat{\omega}_q\tau},$$

which is extremely simple to implement.

3.2 Channel Equalization

Once the channel parameters $\{h_q[l]\}$, $\{\omega_q\}$ are available, we can form

$$u[n] = y[n] - \sum_{l=0}^{L-1} h[n;l]b[n-l].$$

In this section, we will focus on recovering $s[n]$ from $u[n]$.

3.2.1 Linear Frequency Domain Equalizer

First, let us express

$$\begin{aligned} u[n] &= \sum_{l=0}^{L-1} h[n;l]s[n-l] + v[n] && (3.10)\\ &= \sum_{q=1}^{Q} e^{j\omega_q n} \sum_{l=0}^{L-1} h_q[l]s[n-l] + v[n] && (3.11)\\ &= \sum_{q=1}^{Q} e^{j\omega_q n} \left[h_q[n] * s[n]\right] + v[n]. && (3.12) \end{aligned}$$

Its DTFT yields,

$$U(\omega) = \sum_{q=1}^{Q} H_q(\omega - \omega_q)S(\omega - \omega_q) + V(\omega),$$

or equivalently,

$$U(\omega + \omega_1) = H_1(\omega)S(\omega) + \sum_{q=2}^{Q} H_q(\omega - \omega_q + \omega_1)S(\omega - \omega_q + \omega_1) + V(\omega + \omega_1).$$

Since the DTFT is periodic in ω with period 2π, we can set up a system of linear equations to solve for $S(\omega)$ in terms of $U(\omega)$ and $H_q(\omega)$. We assume that a guard interval of length $L-1$ is added after each block of N samples of $s[n]$ to facilitate IBI free block-by-block processing of $y[n]$ [50].

Next, we elaborate on the procedure for the case with $Q = 2$. Let $2\pi b/N_1$ and $2\pi c/N_1$ be the best approximations for ω_1 and $\omega_2 - \omega_1$, respectively, and $N_1 \geq N + L - 1$ is the FFT length (i.e., zero-pad $u[n]$ and $h_q[l]$ to length N_1 prior to taking the FFT). We obtain for $0 \leq k \leq N_1 - 1$,

$$U[k+b] = H_1[k]S[k] + H_2[k-c]S[k-c] + V[k+b],$$

where $U[k] = U[\omega]|_{\omega=2\pi k/N_1}$, and $S[k]$, $H_q[k]$ are similarly defined. Since $U[k]$, $S[k]$, and $H_q[k]$ all have period N_1, we can rewrite the above equation in matrix form as

$$\mathbf{U} = \mathbf{HS} + \mathbf{V},$$

where $\mathbf{H}$ is a sparse $N_1 \times N_1$ matrix with $[H_1[0],\ \ldots,\ H_1[N_1-1]]$ on the main diagonal, $[H_2[0],\ \ldots,\ H_2[N_1-1-c]]$ on the c-th diagonal below the main diagonal, and $[H_2[N_1-c],\ \ldots,\ H_2[N_1-1]]$ on the $[N_1-c]$-th diagonal above the main diagonal; $\mathbf{S} = [S[0],\ \ldots,\ S[N_1-1]]^T$; $\mathbf{U} = [U[b],\ \ldots,\ U[N_1-1],\ U[0],\ \ldots,\ U[b-1]]^T$, and $\mathbf{V}$ is similarly defined as for $\mathbf{U}$. The LS solution is

$$\mathbf{S}_{LS} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{U}.$$

Afterwards, we take the inverse DFT followed by thresholding to produce an estimate for $s[n]$.

The advantage of this equalizer is that it is a linear solution. The drawback however, is that a large $(N_1 \times N_1)$ matrix inversion is involved. Since $\mathbf{H}$ is sparse and block diagonal, numerically efficient solutions may be devised.

3.2.2 Iterative Decision Feedback Equalizer

In Section 3.1.2, we have developed a method for estimating the channel parameters using an FFT-based approach. Since there is also information about $h[n;l]$ contained in the additive "noise" process $u[n]$ (cf. (3.10)), there is room to further improve the $h[n;l]$ estimate. In this subsection, we will describe a joint data - channel estimator which sequentially decodes the symbols $s[n]$ in a decision feedback manner. Once a block of $s[n]$ estimates are obtained, we then go back to refine the channel estimate. With 2-5 iterations, bit error rate (BER) improvements can be observed.

Let us denote the channel parameter estimates during the ith iteration as $h_q^{(i)}[l]$ and $\omega_q^{(i)}$, and the corresponding channel estimate as

$$h^{(i)}[n;l] = \sum_{q=1}^{Q} h_q^{(i)}[l] \exp\{j\omega_q^{(i)} n\}.$$

The initial (i.e., $i = 1$) parameter estimates are obtained as described in Section 3.1.2. Let us form the ith residual process as

$$u^{(i)}[n] = y[n] - \sum_{l=0}^{L-1} h^{(i)}[n;l]b[n-l], \tag{3.13}$$

which should approximate $\sum_{l=0}^{L-1} h[n;l]s[n-l]$ according to (3.10). Denote by $\hat{s}^{(i)}[n]$ the symbol estimate before thresholding, and by $\bar{s}^{(i)}[n]$ the symbol estimate after thresholding, at time n during the ith iteration.

Recall that to enable IBI-free block-by-block processing, we assume that $L-1$ zeros are padded after each block of N samples of $s[n]$ [50]. Therefore, $\bar{s}^{(i)}[n] = 0$ for $-(L-1) \leq n \leq -1$, $\forall i$. Based on (3.10), we first calculate

$$\hat{s}^{(i)}[n] = \frac{u^{(i)}[n] - \sum_{l=1}^{L-1} h^{(i)}[n;l]\bar{s}^{(i)}[n-l]}{h^{(i)}(n;0)}, \tag{3.14}$$

and then threshold it to obtain symbol estimate $\bar{s}^{(i)}[n]$.

Once $\{\bar{s}^{(i)}[n]\}_{n=0}^{N-1}$ are obtained, we form

$$d^{(i+1)}[n] = y[n] - \sum_{l=0}^{L-1} h^{(i)}[n;l]\bar{s}^{(i)}[n-l], \tag{3.15}$$

and use it in place of $y[n]$ in Section 3.1.2, in order to refine the parameter estimates and update the channel estimate to $h^{(i+1)}[n;l]$. Repeating (3.13)-(3.15) 2-5 times usually results in successively improved BER performance.

3.3 Simulations

3.3.1 Frequency Estimation

Let us consider the channel model (2.2) with $h_1[l] = 0.2329 + 0.2312j$, 0, $-0.2362 + 0.2383j$, $0.1465 - 0.2982j$, $h_2[l] = 0.4914 - 0.3534j$, $-0.4431 + 0.3157j$, 0, $0.0834 + 0.0022j$, for $l = 0, 1, 2, 3$, $\omega_1 = 2\pi/120 = 0.0524$ and $\omega_2 = 2\pi/50 = 0.1257$. The pilot sequence $b[n] = 0.8944 \sum_k \delta(n - 4k)$, so $f_q[\tau] = 0.8944 h_q[\tau]$ with $|f_1[\tau]| = 0.2935,\ 0,\ 0.3001,\ 0.2971$, and $|f_2[\tau]| = 0.5414,\ 0.4866,\ 0,\ 0.0746$, for $\tau = 0, 1, 2, 3$. The information sequence $s[n]$ was BPSK $\{\sigma_s, -\sigma_s\}$, and the number of symbols $N = 2,048$. The signal-to-noise ratio defined as SNR $= (\sigma_s^2 + \sigma_p^2)/\sigma_v^2$, was 10 dB. The power allocation factor,

$$\beta = \sigma_p^2/(\sigma_s^2 + \sigma_p^2), \tag{3.16}$$

was 0.2 (i.e., 20% of the transmitted power was allocated to $b[n]$).

Fig. 3.1 shows sequentially from top to bottom, $|Y_0[\alpha]|$, $|Y_1[\alpha]|$, $|Y_2[\alpha]|$, $|Y_3[\alpha]|$, and $|\mathcal{Y}[\alpha]|$. Since $f_1(1) = f_2(2) = 0$, and $|f_2(3)|$ is very small, only one peak is visible from $|Y_1[\alpha]|$, $|Y_2[\alpha]|$, or $|Y_3[\alpha]|$. However, by multiplying, $|\mathcal{Y}[\alpha]| = \prod_{\tau=0}^{3} |Y_\tau[\alpha]|$ clearly reveals the peaks at $4\omega_1 = 0.0668\pi$ and $4\omega_2 = 0.1600\pi$. The estimated frequencies were $\hat{\omega}_1 = 0.0525$, $\hat{\omega}_2 = 0.1257$, which are very close to the true values.

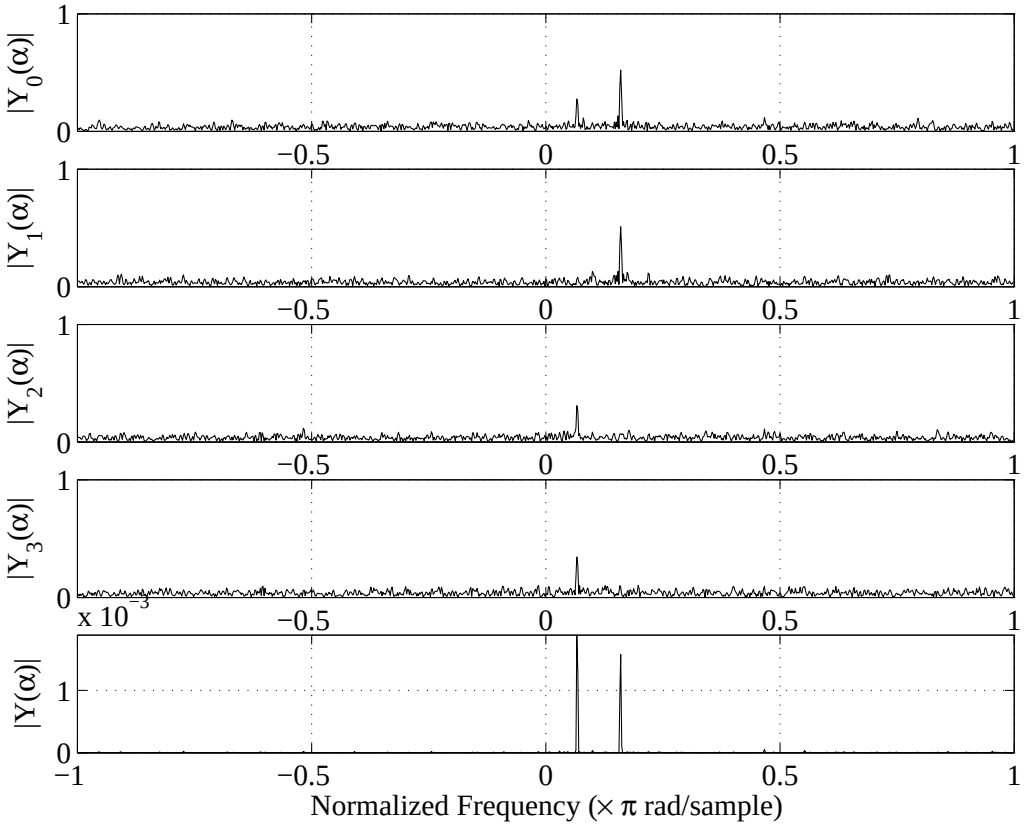


Figure 3.1: From top to bottom: $|Y_0[\alpha]|$, $|Y_1[\alpha]|$, $|Y_2[\alpha]|$, $|Y_3[\alpha]|$, and $|\mathcal{Y}[\alpha]|$.

3.3.2 BER Improvement with Iterations

We focus on the iterative DFE of Section 3.2.2 from now on since it performs better than the linear frequency domain equalizer of Section 3.2.1. Here, $N = 2,048$, $h_1[l] = 0.1790 + 0.1777j$, $0.3263 - 0.3779j$, $-0.1816 + 0.1832j$, $0.1126 - 0.2292j$, $h_2[l] = 0.3777 - 0.2717j$, $-0.3406 + 0.2427j$, $0.3836 - 0.1126j$, $0.0641 + 0.0017j$, $\omega_1 = 2\pi/120$ and $\omega_2 = 2\pi/50$. The pilot $b[n] = \sigma_p \exp\{j\pi/4\} \exp\{-j\pi n^2/4\}$ is periodic in n with period $L = 4$; its DFT $B[k] =$

$2\sigma_p \exp\{j\pi k^2/4\}$ is a chirp as well. We fixed the transmission power at $\sigma_s^2 + \sigma_p^2 = 1$ and the power allocation factor at $\beta = 0.2$ but varied $\sigma_v^2 = N_0/2$ to generate the BER performance as a function of $E_b/N_0 = (\sigma_s^2 + \sigma_p^2)/(2\sigma_v^2)$. Fig. 3.2 shows from top to bottom, BER curves for the DFE with iteration numbers $i = 1, 2, \ldots, 6$, followed by the performance resulting from a perfectly known channel. We observe that iteration improves the BER performance, especially at high E_b/N_0 values. If we define the channel mean square error as

$$E_h^{(i)} = N^{-1} \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} |h^{(i)}[n;l] - h[n;l]|^2,$$

we can observe (plot not shown here) that as i increases, improvements in $E_h^{(i)}$ are well correlated with decreases in BER as shown in Fig. 3.2. This implies that symbol decoding and channel estimation are mutually beneficial processes. However, there is diminishing return in carrying out more than 5 iterations. In Fig. 3.2, the BER curve corresponding to $i = 5$ is no more than 1.5 dB apart from the case with perfect channel knowledge.

3.3.3 BER and the Power Allocation Factor

For fixed transmission power $\sigma_s^2 + \sigma_p^2$, we have to weigh σ_s^2 and σ_p^2 carefully in order to obtain good BER. A high σ_p^2 can produce a good channel estimate, but little power is left to transmit $s[n]$. On the other hand, a high σ_s^2 does not help if the channel state information is inaccurate. In Fig. 3.3, we show the BER after 5 iterations, for varying values of β. The same parameters as in Section 3.3.2 are used here. In this example, $\beta = 0.2$ offers the best compromise between channel sounding and symbol transmission. Theoretical analysis of the optimum β is beyond the scope of this work, but the techniques of [10] can be of interest.

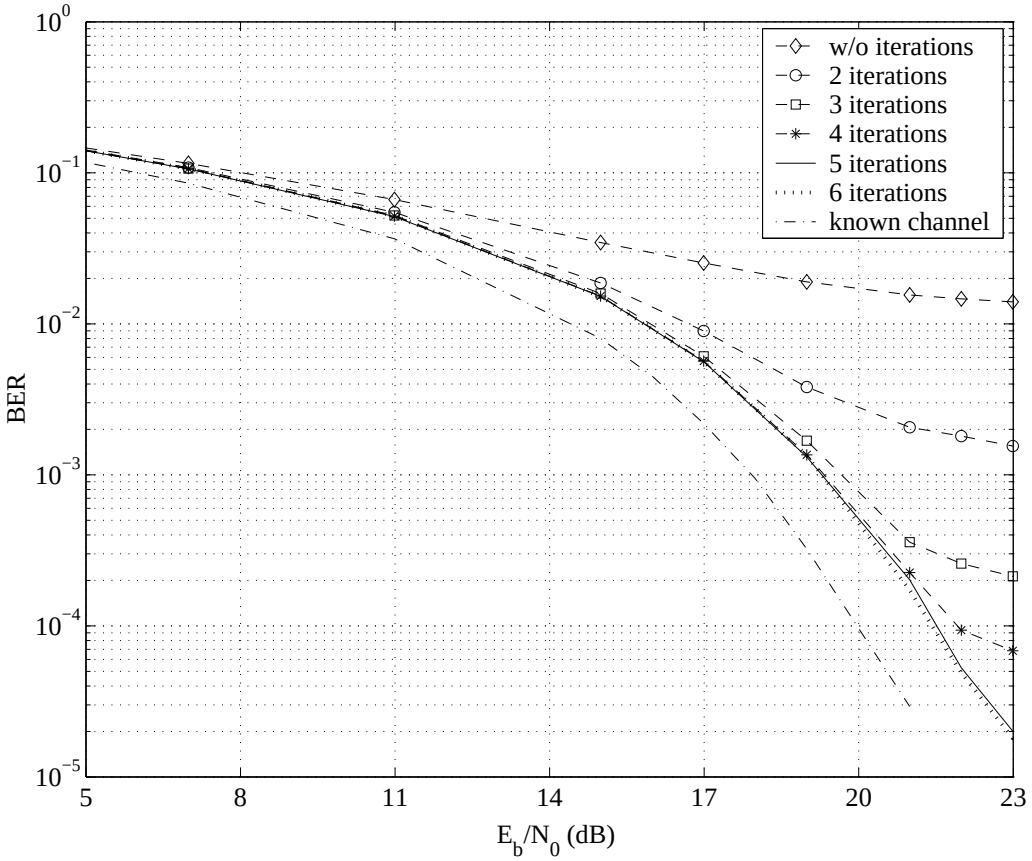


Figure 3.2: BER vs. the number of iterations.

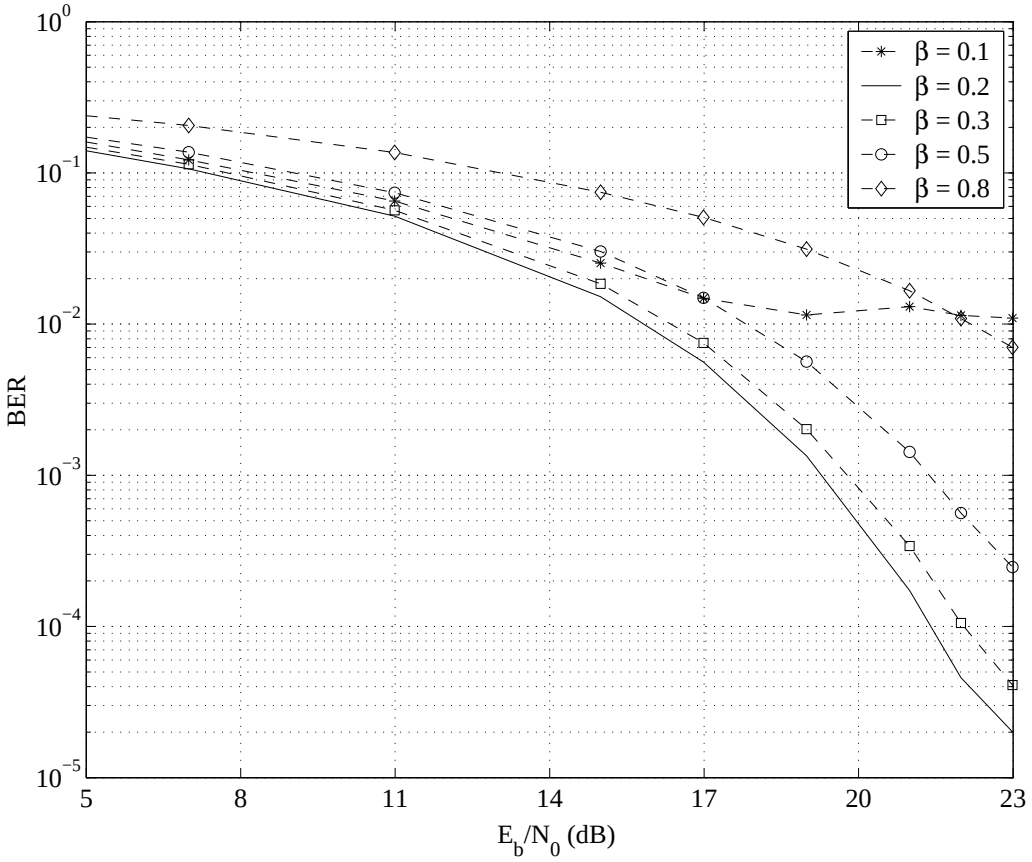


Figure 3.3: BER performance for varying β.

CHAPTER 4

IMPROVING POWER EFFICIENCY BY PAR REDUCTION

In this chapter, we propose a dynamic SLM technique, which greatly reduces the computational requirement of SLM without sacrificing its PAR reducing capability. We propose three blind SLM techniques, one for OFDM and two for forward link CDMA.

4.1 Selected Mapping with Thresholding and Clipping

In this section, we describe a dynamic SLM method with OFDM as the model system. We emphasize however that this technique is readily applicable to CDMA as well.

From (2.6), we can see that the worst possible PAR of an OFDM signal is N, which occurs at $t = 0$ when $X[k]$ are all the same. For example, when $N = 128$, the worst possible PAR is $10 \log_{10}(128) = 21$ dB. To amplify this signal absolutely without any distortion; i.e., to avoid any possibility of clipping, we need to position the highest possible peak power at $\mathcal{P}_{\text{max1}}$ in Fig. 1(a). Under this arrangement, the average power $\mathcal{P}_{\text{av1}}$ will be very low and the PA efficiency will be depressed to less than 1%. In practice, a PA is expected to provide a certain level of power efficiency, which means that for a given PA and biasing conditions, the average input power $\mathcal{P}_{\text{av}}$ has to be above a certain amount. This also requires the input signal PAR to be less than a threshold γ_0. The concept of PAR thresholding was explored in [27] for the PTS technique.

In [78], it was shown that as long as $E[e^{j\phi^{(m)}[k]}] = 0$, $\forall\ k, m$, the time-domain signals $z^{(m)}(t)$ and $z^{(l)}(t)$ will be asymptotically independent for $m \neq l$. This means that for a

large N, we can obtain the CCDF of the SLM-OFDM signal $z^{(\bar{m})}(t)$ as follows:

$$\begin{aligned}
\Pr\left\{\mathrm{PAR}(z^{(\bar{m})}(t)) > \gamma\right\} &= \Pr\left\{\min_{1\le m\le M} \mathrm{PAR}(z^{(m)}(t)) > \gamma\right\} \\
&= \Pr\left\{\mathrm{PAR}(z^{(m)}(t)) > \gamma,\ \forall\ 1 \le m \le M\right\} \\
&= \left[\Pr\left\{\mathrm{PAR}(x(t)) > \gamma\right\}\right]^M \\
&= \left[1 - a\right]^M, \qquad (4.1)
\end{aligned}$$

where $a = \exp\left\{-e^{-\gamma} N \sqrt{\frac{\pi}{3}\ln N}\right\}$ (c.f. (2.8)).

We make the following remarks regarding the "conventional" SLM described above:

1. SLM aims at minimizing the PAR per OFDM block by carrying out all M mappings. Even if the first few mappings have already managed to reduce the PAR to below a certain threshold γ_0, the SLM scheme still continues to seek further reduction of the PAR.

2. For given N and γ_0 values, equation (4.1) shows that even after all M mappings are tried out, there is still a non-zero probability that the SLM method fails to meet the PAR goal; i.e., the resulting $\mathrm{PAR}(z^{(\bar{m})}(t)) > \gamma_0$. When that happens, $z^{(\bar{m})}(t)$ will need to be clipped to meet the peak power and average power constraints.

3. For given N and M values and clipping probability $p = \Pr\{\mathrm{PAR}(z^{(\bar{m})}(t)) > \gamma_0\}$, we can find from (4.1), the corresponding PAR threshold

$$\gamma_0 = \ln(N\sqrt{\frac{\pi}{3}\ln N}) - \ln(\ln\frac{1}{1-p^{1/M}}). \qquad (4.2)$$

We investigate next, a modified SLM technique which incorporates the above PAR thresholding and clipping considerations.

4.1.1 Dynamic SLM Algorithm

Our objective here is to apply SLM, but to stop trying as soon as the PAR threshold γ_0 is met, with the constraint that the number of trials is no more than M (including the original OFDM signal). For example, for $N = 128$, $\gamma_0 = 5.6234$ (7.5 dB), there is a 35.27% chance that no phase rotations are ever needed ($M = 1$); there is a 58.09% chance that only one

Table 4.1: Dynamic SLM algorithm.

Step 0	Set $\bar{m} = m = 1$.
Step 1	Form $z^{(m)}(t)$, and compute $\text{PAR}(z^{(m)}(t))$.
Step 2	If $\text{PAR}(z^{(m)}(t)) \leq \gamma_0$, then continue to Step 3; else go to Step 4.
Step 3	Set $\bar{m} = m$ and $\tilde{z}(t) = z^{(\bar{m})}(t)$, and go to Step 7.
Step 4	If $\text{PAR}(z^{(m)}(t)) < \text{PAR}(z^{(\bar{m})}(t))$, then go to Step 4a; else go to Step 4b.
	Step 4a Set $\bar{m} = m$.
	Step 4b $m = m + 1$.
Step 5	If $m > M$, then go to Step 6; else go to Step 1.
Step 6	Clip $z^{(\bar{m})}(t)$ to form $\tilde{z}(t) = \begin{cases} z^{(\bar{m})}(t), & \text{if } \lvert z^{(\bar{m})}(t)\rvert \leq A, \\ A\ \exp\{j\angle z^{(\bar{m})}(t)\}, & \text{otherwise,} \end{cases}$
Step 7	Transmit $\tilde{z}(t)$.

phase mapping suffices ($M = 2$); etc. Our strategy is "to do only what is necessary" is order to save computational resources. With this approach, it is possible to increase the data rate as compared with the conventional SLM, by resorting to a two-buffer, dynamic scheme described in [53]. As mentioned before, there is always the possibility that even after all M trials, SLM still fails to meet the PAR goal γ_0. In that case, $z^{(\bar{m})}(t)$ is clipped to become $\tilde{z}(t)$, which has maximum amplitude $A = \sqrt{\mathcal{P}_{\text{av}}\ \gamma_0}$ (the clipping level). As long as the clipping probability (4.1) evaluated at γ_0 is small (e.g., 10^{-3}), there will be negligible amount of spectral regrowth or BER increase.

The step-by-step algorithm for the proposed dynamic SLM; i.e., SLM with thresholding and clipping (SLMTC) technique is described in Table 4.1. Fig. 4.1 shows the algorithm in the flow chart format.

4.1.2 Performance Analysis

We analyze here, the CCDF expression for the SLMTC signal $\tilde{z}(t)$ obtained as described in the previous subsection. Denote by $z^{(\bar{m})}(t)$ the signal after SLM with thresholding, which is *not* to be confused with the $z^{(\bar{m})}(t)$ notation used in the conventional SLM (c.f. (2.10)). If $\gamma \leq \gamma_0$, the event $\text{PAR}(\tilde{z}(t)) \leq \gamma$ is equivalent to the event $\text{PAR}(z^{(\bar{m})}(t)) \leq \gamma$, which in turn is equivalent to the event

$$\exists\ 1 \leq d \leq M, \text{ such that } \text{PAR}(z^{(d)}(t)) \leq \gamma \text{ and } \left\{\text{PAR}(z^{(l)}(t) > \gamma_0)\right\}_{l=1}^{d-1}. \tag{4.3}$$

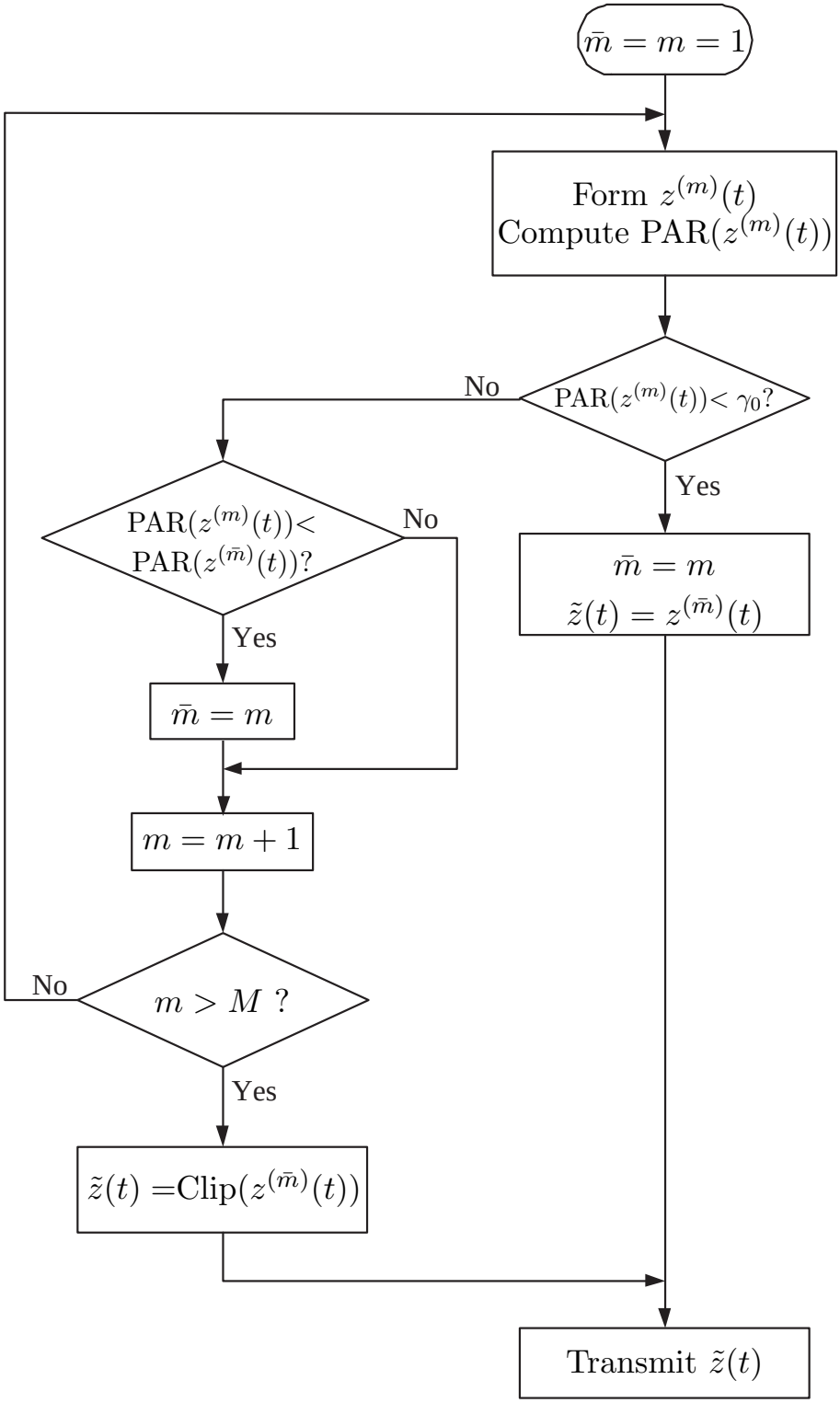


Figure 4.1: Flow chart of the dynamic SLM algorithm.

By recalling (2.8), we obtain

$$\begin{aligned}\Pr\{\mathrm{PAR}(\tilde{z}(t)) \le \gamma\} &= \sum_{d=1}^{M} \Pr\{\mathrm{PAR}(z^{(d)}(t)) \le \gamma\} \prod_{l=1}^{d-1} \Pr\{\mathrm{PAR}(z^{(l)}(t)) > \gamma_0\} \\ &= \sum_{d=1}^{M} a\ (1-a_0)^{d-1} \\ &= \frac{a}{a_0}(1-(1-a_0)^M), \quad \text{for} \ \ \gamma \le \gamma_0, \end{aligned} \tag{4.4}$$

where $a_0 = \exp\left\{-e^{-\gamma_0} N \sqrt{\frac{\pi}{3}\ln N}\right\}$.

Obviously due to clipping,

$$\Pr\{\mathrm{PAR}(\tilde{z}(t)) > \gamma\} = 0, \quad \text{for} \ \ \gamma > \gamma_0. \tag{4.5}$$

Combining (4.4) and (4.5), we find the CCDF of the PAR for the proposed SLMTC method:

$$\Pr\{\mathrm{PAR}(\tilde{z}(t)) > \gamma\} = \begin{cases} 1 - \frac{a}{a_0}\left[1-(1-a_0)^M\right], & \gamma \le \gamma_0 \\ 0, & \gamma > \gamma_0, \end{cases} \tag{4.6}$$

with $a = \exp\left\{-e^{-\gamma} N \sqrt{\frac{\pi}{3}\ln N}\right\}$ and $a_0 = \exp\left\{-e^{-\gamma_0} N \sqrt{\frac{\pi}{3}\ln N}\right\}$.

4.1.3 Simulations

Assume that the number of sub-carriers is $N = 128$, the maximum number of phase rotations is $M = 16$, and the PAR threshold is $\gamma_0 = 7.5$ dB. The frequency-domain OFDM sub-symbols were drawn independently from a QPSK constellation, and 10^6 Monte Carlo runs were performed.

Fig. 4.2 shows the empirical CCDFs (solid lines) of $\mathrm{PAR}(x(t))$ (OFDM), $\mathrm{PAR}(z^{(\bar{m})}(t))$ (SLM), and $\mathrm{PAR}(\tilde{z}(t))$ (SLMTC), along with the corresponding theoretical CCDFs (dash-dotted lines) evaluated using (2.8), (4.1), and (4.6), respectively. The empirical CCDFs of the continuous-time PAR in (2.3) were obtained by evaluating the discrete-time PAR of the 4-time oversampled OFDM signal [62]. It is evident from Fig. 4.2 that the theoretical and the empirical CCDFs agreed very well. We observe that when $M = 16$, the proposed algorithm achieved 3.5 dB of PAR reduction at the CCDF level of 10^{-3}. Indeed, if we substitute $N = 128$, $M = 16$, and $p = 10^{-3}$ into (4.2), we obtain $\gamma_0 = 5.6178 = 7.4957$ (dB).

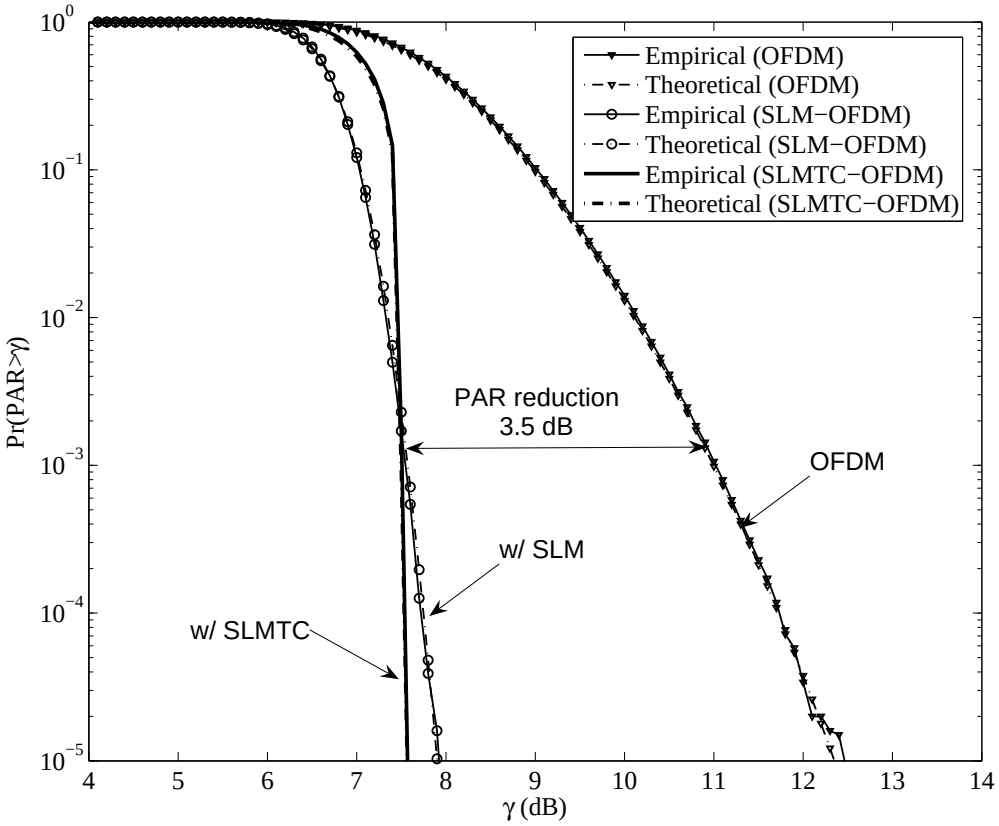


Figure 4.2: CCDF of the PAR of the OFDM signal and OFDM with SLM or SLMTC.

We observe from Fig. 4.2 that the CCDF curves for SLM and SLMTC cross over at γ_0, and that for $\gamma < \gamma_0$, SLMTC has less PAR reducing capability than SLM. This is completely expected since by design, SLMTC generally uses fewer mappings and consumes less computational resources than SLM. Unless one pursues block-by-block adaptive biasing or linear scaling [47] approaches, any PAR value lower than the required γ_0 does not necessarily lead to additional power savings. We regard SLMTC as a lower cost alternative to SLM. As we mentioned in Section 4.1.1, the resources savings from the PAR thresholding can be harvested using a buffered dynamic processing scheme [53], which results in a smaller transmission latency than SLM and thus permits a higher data rate.

4.1.4 Testbed Experiments

We have conducted testbed experiments on two different PAs to demonstrate our approach. Our goal is to show that for the same PA, it is possible to boost the average transmit power through PAR reduction and PA linearization, while keeping the ACPR unchanged. On the other hand, without PAR reduction or PA linearization, the increase in the average power would come at a cost of spectral regrowth; i.e., larger ACPR, which may violate the spectral regulations.

Our testbed consists of a high-speed digital I/O system, a digital-to-analog converter (DAC), RF transmit and receive chains, a device under test (DUT), and an analog-to-digital converter (ADC). The high-speed digital I/O system has 150 million samples per second (MSPS) 16-bit digital input/output capability. In the transmission mode, the digital I/O system first generates baseband data, applies the SLMTC algorithm, predistorts it, and then digitally up-converts the signal to an intermediate frequency (IF) of 30 MHz, and finally sends out the 14-bit data stream to the DAC at a sampling rate of 120 MSPS. Superheterodyne up-conversion and down-conversion chains are used to convert the digital IF signal to and from the carrier frequency. The DUTs are respectively, a 1 W handset PA and a 45 W base-station PA. In the acquisition mode, the digital I/O system acquires 12-bit digital IF data at the sampling rate of 120 MSPS from the ADC. The received baseband data $y[n]$ is obtained by converting the PA output to baseband and removing the time delay between the input and the output of the digital I/O system. Since the signal is modulated in the digital domain, any in-phase and quadrature imbalance problem in the quadrature modulator is obviated.

4.1.4.1 Experiment on a 1 W handset PA

In this experiment, the DUT is the 1 W handset PA. The input is an OFDM signal centered at 836 MHz with a 1.25 MHz bandwidth and 128 subcarriers.

We measured the power spectral density (PSD) of the PA output using a spectrum analyzer. ACPR was measured as the ratio between the average power in the adjacent channel and the average power in the main channel, both over a 30 KHz bandwidth [16] [81].

The requirement was to keep the ACPR below -50 dBc. Fig. 4.3 shows the PSDs of the PA output when (a) the input was backed-off just enough to meet the ACPR requirement; (b) a memoryless polynomial PD (i.e., $Q = 0,\ K = 5$ in (2.5)) was applied, and the amount of input back-off was reduced; (c) both SLMTC ($M = 16,\ \gamma_0 = 7.5$ dB) and the memoryless polynomial PD were applied, requiring even less input back-off. By comparing curves (a) and (b) in Fig. 4.3, we see that the average output power in the main channel increased by 6 dB thanks to the use of the PD and the resulting reduction in back-off. Moreover, with the SLMTC PAR reduction technique, we were able to boost the average output power by another 3 dB without introducing any spectral regrowth (compare lines (b) and (c)). Therefore, we have achieved a total of 9 dB increase in the average output power of the PA by the combination of PAR reduction and predistortion linearization.

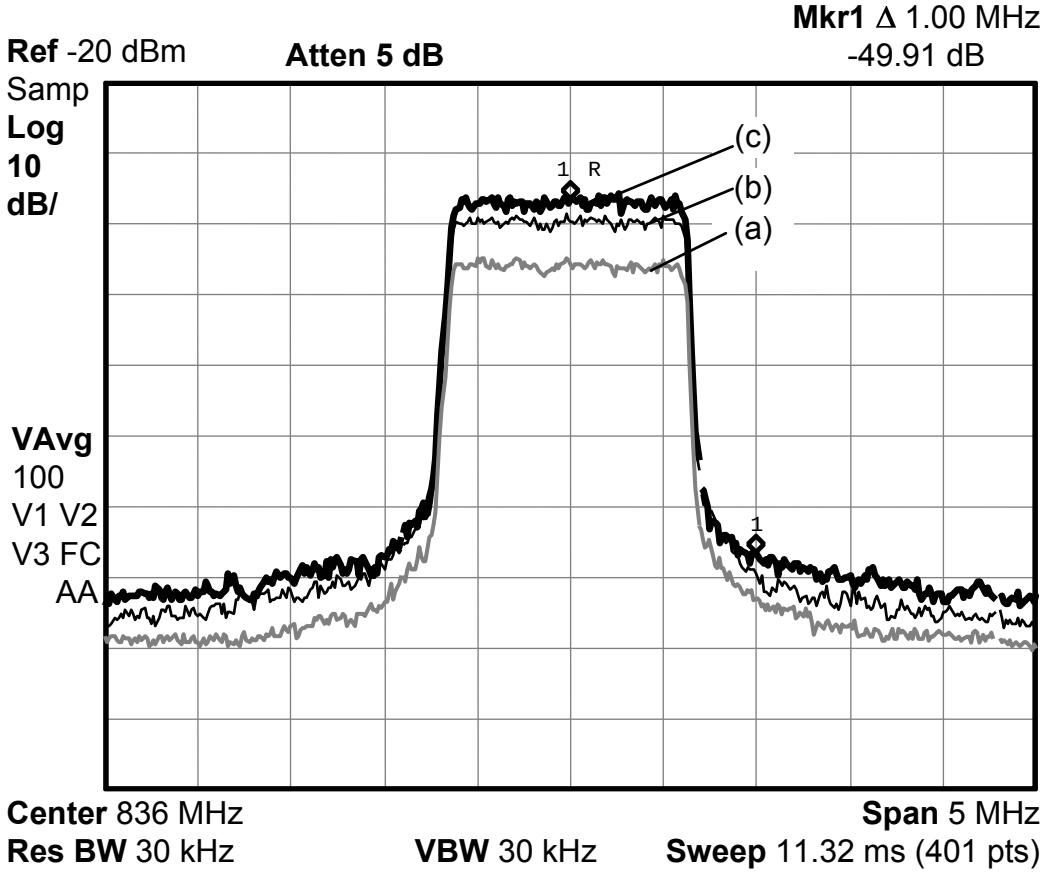


Figure 4.3: Power spectral density measurements at the output of the 1 W handset PA when (a) the input was backed-off, (b) a memoryless polynomial PD ($Q = 0,\ K = 5$) was applied, and (c) both SLMTC ($M = 16,\ \gamma_0 = 7.5$ dB) and the memoryless polynomial PD ($Q = 0,\ K = 5$) were applied.

4.1.4.2 Experiment on a 45 W base-station PA

In this experiment, the DUT is the 45 W base-station PA. The input is an OFDM signal centered at 881 MHz with a 2.5 MHz bandwidth and 128 subcarriers.

For the 45 W PA, the requirement was to keep the ACPR below -45 dBc. Fig. 4.4 shows the PSDs of the PA output when (a) the input was backed-off just enough to meet the ACPR specification; (b) a memory polynomial PD (i.e., $Q = 5,\ K = 5$ in (2.5)) was applied; (c) both SLMTC ($M = 16,\ \gamma_0 = 7.5$ dB) and the memory polynomial PD were applied. From Fig. 4.4, we can see that the average output power was increased by 11 dB through the combination of PAR reduction and predistortion linearization. Through experimentation, we have found that this high power amplifier had significant memory effects and memoryless predistortion was not as effective as the memory polynomial predistortion demonstrated here.

In this section, we discussed a thresholding and clipping technique to reduce the computational resource requirements of selected mapping (SLM). A closed form CCDF expression was derived and was shown to agree with the empirical results very well. We argue that PAR reduction alters the statistical characteristics of the signal, and PA linearization changes the effective response of the nonlinear PA, thus both techniques can be applied independently. Using testbed experiments, we demonstrated the effectiveness of our technique as significant increase in the average output power without exceeding the spectral emission limits. Our analysis uses OFDM as the model system, but the idea of thresholding and clipping applies to other systems characterized by high PAR values as well.

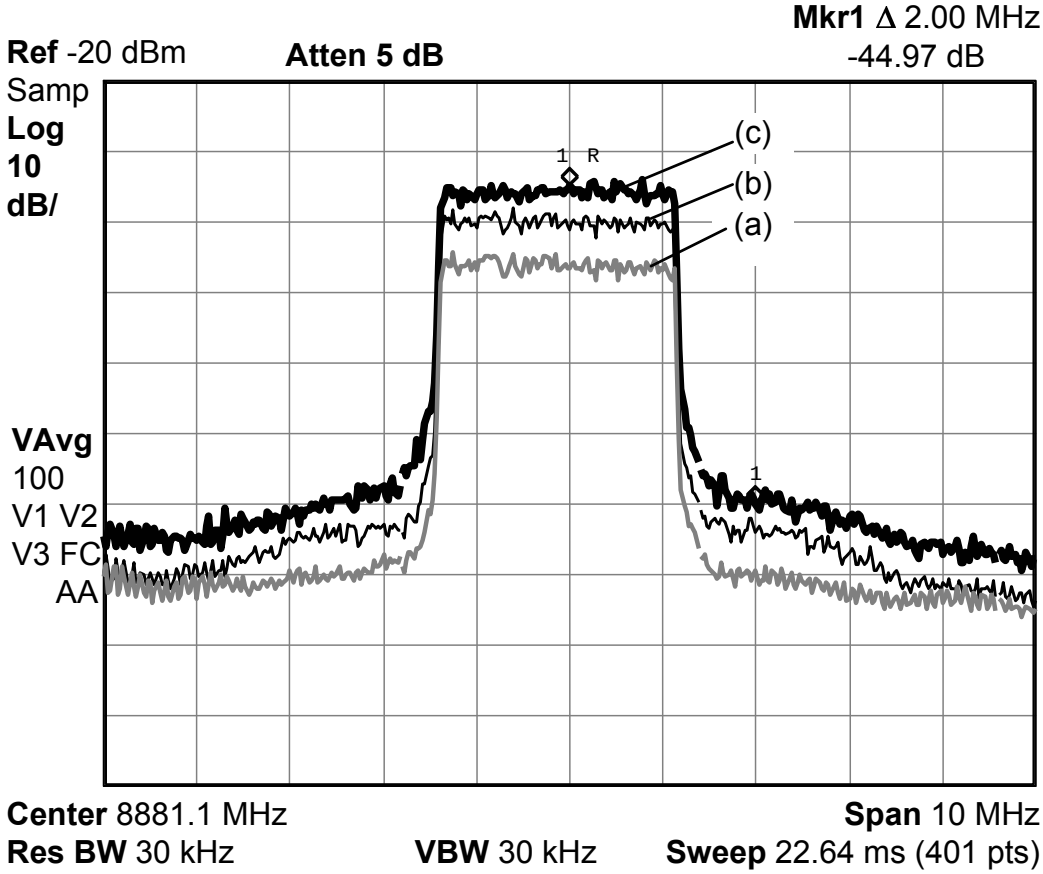


Figure 4.4: Power spectral density measurements at the output of the 45 W base-station PA when (a) the input was backed-off, (b) a memory polynomial PD ($Q = 5,\ K = 5$) was applied, and (c) both SLMTC ($M = 16,\ \gamma_0 = 7.5$ dB) and the memory polynomial PD ($Q = 5,\ K = 5$) were applied.

4.2 Blind SLM for OFDM

In OFDM, the CSI can be acquired by modulating pilot tones onto predetermined subcarriers; this is called pilot tone assisted modulation (PTAM) [46, 50]. We propose next a blind SLM scheme[1] that combines the merits of PTAM and SLM. In our joint channel estimation and PAR reduction algorithm, instead of fixing the pilot tone locations as in conventional PTAM, we propose to try different pilot tone locations, and synchronize the movement of the pilot tones with the choice of the phase rotation sequence. The pilot tone / phase sequence combination that results in the lowest PAR of the time-domain signal is used for transmission. However, the optimum index is not transmitted as side information in order to maintain the information rate. At the receiver, by taking advantage of the disparity between the pilot tone and information signal powers, we can blindly detect the optimum index by resorting to simple frequency-domain averages. The concept of joint channel estimation and PAR reduction was also explored in [21] – the "diversity" offered by the pilot phase (as opposed to the pilot location) was exploited and the transmission of side information was assumed in [21].

We first review the PTAM technique for OFDM in Section 4.2.1, and describe our proposed blind selected pilot tone modulation (BSPTM) technique in Section 4.2.2. In Section 4.2.3, we use computer simulations to demonstrate the impressive PAR reducing capability of the proposed algorithm and its robust BER performance over frequency selective fading channels.

4.2.1 PTAM for OFDM

In an OFDM transmission system with PTAM, P pilot tones are inserted in the frequency domain in order to acquire the CSI; $P \geq L$ is assumed where L is the length of the finite impulse response (FIR) channel. The transmitted frequency domain signal can be described

[1]The blind SLM technique described in this section is protected by the U.S. Utility Patent Application: "Crest Factor Reduction in OFDM with Blind Selected Pilot Tone Modulation," by N. Chen and G. Tong Zhou, filed on March 15, 2006.

by

$$X[k] = \begin{cases} B[k] \text{ for } k \in \mathbf{\Omega}_0, \\ S[k] \text{ for } k \in \mathbf{\Omega}_0^{\perp}, \end{cases} \tag{4.7}$$

where $\mathbf{\Omega}_0$ is the set of the P pilot tone indices in ascending order, $\mathbf{\Omega}_0^{\perp}$ denotes the complement of $\mathbf{\Omega}_0$ (i.e., the set of $N-P$ data indices in ascending order), $\{B[k]\}_{k\in\mathbf{\Omega}_0}$ are the pilot tones, and $\{S[k]\}_{k\in\mathbf{\Omega}_0^{\perp}}$ are the frequency-domain information sub-symbols.

According to [50], the optimal way to place the pilot tones is to modulate $P = L$ pilot tones with equal power onto equi-spaced sub-carriers. For simplicity, let us assume that the number of sub-carriers N is an integer multiple of P; i.e., $R = N/P$ is an integer. Next define a set of P equi-spaced pilot tone indices as

$$\mathbf{\Omega}_0 \triangleq \Big\{ k_i \;\Big|\; k_i = iR + \theta_0, 0 \le i \le P-1, 0 \le \theta_0 \le R-1 \Big\}, \tag{4.8}$$

which can be characterized by θ_0 alone.

At the receiver, after removing the cyclic prefix and taking the discrete Fourier transform (DFT), we obtain a set of N linear equations in the frequency domain

$$Y[k] = X[k]H[k] + V[k], \tag{4.9}$$

where

$$Y[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} y[n] e^{-j2\pi kn/N}$$

and

$$V[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} v[n] e^{-j2\pi kn/N}$$

are the normalized DFT of the received signal $y[n]$ (after the removal of the cyclic prefix) and the zero-mean additive noise $v[n]$, respectively, and

$$H[k] = \sum_{n=0}^{L-1} h[n] e^{-j2\pi kn/N}$$

is the frequency response of the composite channel (the convolution of the transmit filter, the frequency selective channel, and the receive filter).

Since $X[k] = B[k]$ for $k \in \mathbf{\Omega}_0$, we obtain from (4.9) an estimate of $H[k]$ at P points of $\mathbf{\Omega}_0$:

$$\hat{H}[k] = \frac{Y[k]}{B[k]}, \quad k \in \mathbf{\Omega}_0. \tag{4.10}$$

Since $H[k]$ is constrained by P parameters $\{h[n]\}_{n=0}^{P-1}$, we can then estimate $H[k]$ at any k.

Afterwards, the information sub-symbols can be estimated as

$$\hat{S}[k] = \frac{Y[k]}{\hat{H}[k]}, \quad k \in \mathbf{\Omega}_0^{\perp}, \tag{4.11}$$

which are then decoded to yield the $\bar{S}[k]$ estimates belonging to the symbol constellation.

4.2.2 Blind Selected Pilot Tone Modulation

In this section, we describe our proposed BSPTM technique which is a combination of channel sounding and effective PAR reduction, at a low computational cost.

4.2.2.1 Disparity in the Pilot and Information Signal Powers

We first point out an interesting feature of PTAM: *the pilot tones generally have stronger average power than the information sub-symbols* – this forms the basis of our BSPTM approach.

Let us denote by β, the power allocation factor, which is the ratio between the total power allocated to the pilot tones and the total transmitted power; i.e.,

$$\beta = \frac{P\sigma_p^2}{P\sigma_p^2 + (N-P)\sigma_s^2}. \tag{4.12}$$

where $\sigma_p^2 = \frac{1}{P}\sum_{k\in\mathbf{\Omega}_0} |B[k]|^2$ is the average power of the pilot tones and σ_s^2 is the variance of $S[k]$; i.e., $\sigma_s^2 = E[|S[k]|^2]$. Please note that σ_p^2 and β for PTAM are defined in the frequency domain, whereas the corresponding quantities in (3.16) for superimposed training are defined in time domain.

When the pilot tones are equi-powered; i.e., $|B[k]|^2 = \sigma_p^2, \ \forall\ k \in \mathbf{\Omega}_0$, the optimal β was determined in [50] as

$$\beta = 1 - \frac{1}{1 + \sqrt{1/(\frac{N}{P} - 1)}}, \tag{4.13}$$

by minimizing the MSE of the source estimates $\hat{S}[k],\ k \in \mathbf{\Omega}_0^{\perp}$. Combining (4.12) and (4.13), we find

$$\frac{\sigma_p^2}{\sigma_s^2} = \sqrt{\frac{N}{P} - 1}, \tag{4.14}$$

which depends on N/P only. Since $N \gg P$, the pilot tones have much stronger power than the information sub-symbols. For example, for $P \leq 16$ and $N \geq 160$, (4.14) gives rise to $\sigma_p^2/\sigma_s^2 \geq 3$. On the other hand, if $P \leq 8$ and $N \geq 296$, we have $\sigma_p^2/\sigma_s^2 \geq 6$. Both are realistic scenarios.

We will describe later how the $\sigma_p^2/\sigma_s^2 \gg 1$ relationship helps us to detect the pilot tone location parameter θ_0.

4.2.2.2 PAR Reduction by BSPTM

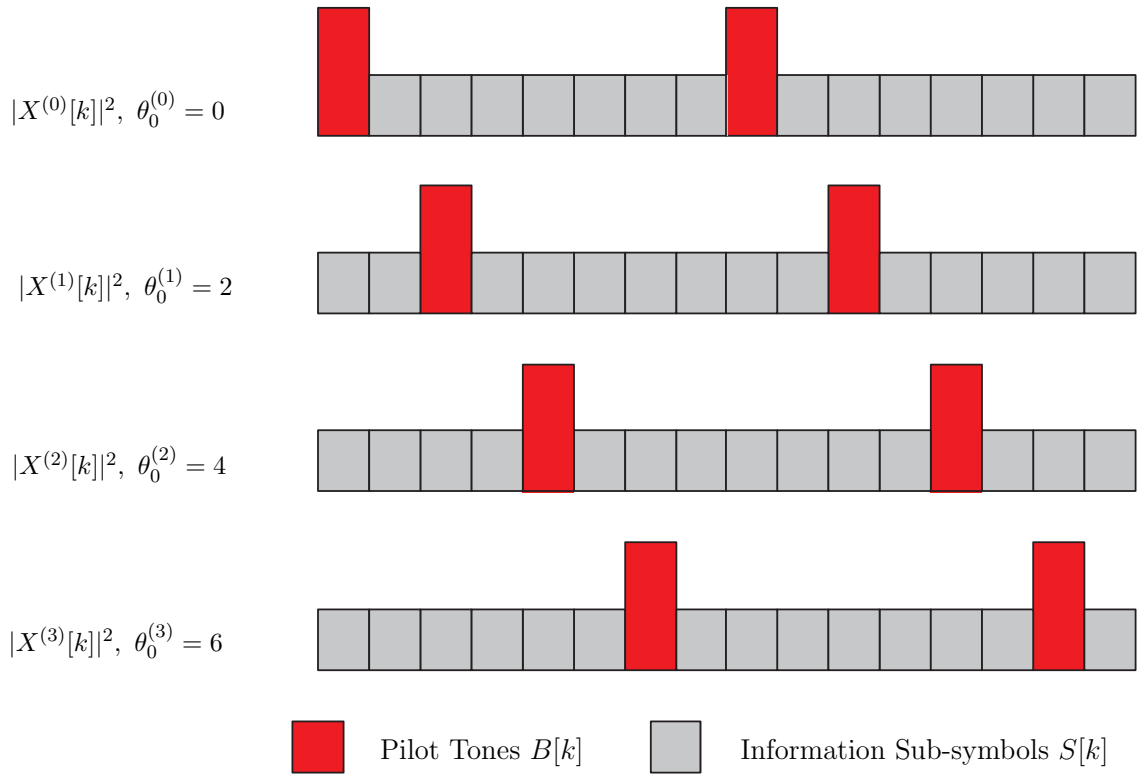


Figure 4.5: An example to illustrate some possible scenarios for $X^{(m)}[k]$ with $N = 16,\ P = 2,\ M = 4$.

According to [50], as long as the pilot tones are equi-powered and equi-spaced and the additive noise is white, channel estimation performance is not affected by the choice of θ_0. Therefore, instead of using a pre-selected θ_0, we can try different frequency shifts $\theta_0^{(m)}$ for

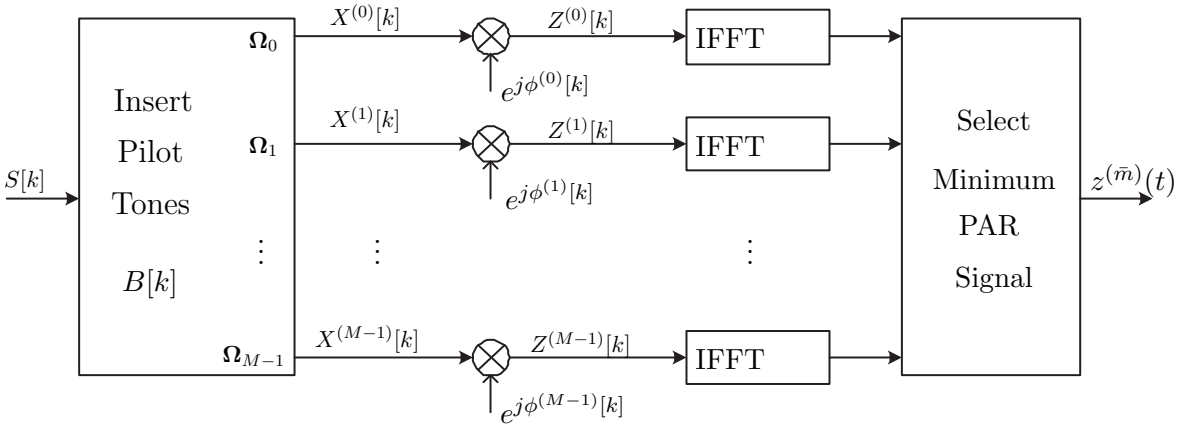


Figure 4.6: Blind selective pilot tone modulation (BSPTM) OFDM transmission. $\mathbf{\Omega}_m$ indicates the mth set of pilot tone locations and $\phi_k^{(m)}$ is the mth phase rotation sequence.

the pilot tones. The novelty of our approach is to tie the location of the pilot tones to the different phase rotation sequences. This enables PAR reduction without the transmission of side information.

Recall that we use m to index the rows of the phase rotation table. Let us use the same m to index the M candidate frequency shifts for the pilot tones; i.e.,

$$\mathbf{\Theta} \triangleq \{\theta_0^{(0)},\ \theta_0^{(1)},\ \ldots,\ \theta_0^{(m)}, \ldots\ \ldots,\ \theta_0^{(M-1)}\}. \tag{4.15}$$

The maximum number of distinct pilot tone locations is $R = N/P$, in which case $\{\theta_0^{(0)} = 0,\ \theta_0^{(1)} = 1,\ \ldots,\ \theta_0^{(R-1)} = R-1\}$. However, since R can be quite large and for practical reasons, we do not usually need M greater than, say 8 (see also Section 4.2.3.1), there is some flexibility in designating $\mathbf{\Theta}$. For example, if $R = 8$ and $M = 4$, we can have $\{\theta_0^{(0)} = 0,\ \theta_0^{(1)} = 2,\ \theta_0^{(2)} = 4,\ \theta_0^{(3)} = 6\}$ (see Fig. 4.5) or $\{\theta_0^{(0)} = 0,\ \theta_0^{(1)} = 1,\ \theta_0^{(2)} = 3,\ \theta_0^{(3)} = 7\}$, and so on. We suggest for the M delays to be equi-spaced in order to minimize the detection error in $\bar{m}$; i.e., the possible pilot shifts are $\{\theta_0^{(0)} = 0,\ \theta_0^{(1)} = R/M,\ \ldots,\ \theta_0^{(M-1)} = \frac{M-1}{M}R\}$. In addition, both the transmitter and the receiver should use the same convention for $\mathbf{\Theta}$.

In BSPTM, the mth PTAM-OFDM signal is given by

$$X^{(m)}[k] = \begin{cases} B[k], & k \in \mathbf{\Omega}_m, \\ S[k], & k \in \mathbf{\Omega}_m^{\perp}, \end{cases}, \quad 0 \le k \le N-1,\ 0 \le m \le M-1, \tag{4.16}$$

where $\mathbf{\Omega}_m$ is characterized by $\theta_0^{(m)}$ similar to the way that $\mathbf{\Omega}_0$ is characterized by θ_0.

Next perform the phase rotations,

$$Z^{(m)}[k] = X^{(m)}[k]e^{j\phi^{(m)}[k]}. \tag{4.17}$$

Similar to SLM, $z^{(m)}(t)$ and $\mathrm{PAR}\left(z^{(m)}(t)\right)$ are evaluated and $z^{(\bar{m})}(t)$, which has the lowest PAR among $\{z^{(m)}(t)\}$, is transmitted. In other words, the optimum pilot tone location - phase sequence index is

$$\bar{m} = \underset{0\le m\le M-1}{\operatorname{argmin}} \left\{\mathrm{PAR}\left(z^{(m)}(t)\right)\right\}. \tag{4.18}$$

4.2.2.3 Performance Analysis

It has been shown in [72] that

$$\Pr\{\mathrm{PAR}(x(t)) > \gamma\} = 1 - e^{-e^{-\gamma}N\sqrt{\frac{\tilde{\lambda}}{N}\log N}}, \tag{4.19}$$

where $\tilde{\lambda} = \lambda_2 - \lambda_1^2$ and

$$\begin{aligned} \lambda_1 &= \frac{\sum_{k=0}^{N-1}\frac{2\pi k}{N}g[k]}{\sum_{k=0}^{N-1}g[k]}, \\ \lambda_2 &= \frac{\sum_{k=0}^{N-1}(\frac{2\pi k}{N})^2 g[k]}{\sum_{k=0}^{N-1}g[k]}, \end{aligned} \tag{4.20}$$

and $g[k]$ is the amount of power allocated to the kth subcarrier. In the PTAM-OFDM system, we have

$$g[k] = \begin{cases} \sigma_p^2, & k \in \mathbf{\Omega}_0 \\ \sigma_s^2, & k \in \mathbf{\Omega}_0^{\perp}. \end{cases} \tag{4.21}$$

Substituting (4.21) into (4.20), we obtain $\lambda_1 = \pi(1-\beta/P)$, $\lambda_2 \approx 4\pi^2/3$ (when $N \gg P$), and $\tilde{\lambda} = 4\pi^2/3 - \pi^2(1-\beta/P)^2$ for PTAM-OFDM.

Based on (4.19), we derive the CCDF of $\mathrm{PAR}(z^{\bar{m}}(t))$ as follows.

From (2.6), the mth time-domain candidate OFDM signal is given by

$$z^{(m)}(t) = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} Z^{(m)}[k]\; e^{j2\pi kt/T_s}, \quad 0 \le t \le T_s, \quad 0 \le m \le M-1. \tag{4.22}$$

We show here that $z^{(m)}(t)$ and $z^{(n)}(t)$ are asymptotically independent processes if $m \neq n$. Since $Z^{(m)}[k] = X^{(m)}[k]e^{j\phi^{(m)}[k]}$ (c.f. (4.17)), and $X^{(m)}[k]$ and $\phi^{(m)}[k]$ are mutually

independent, we infer from $E[e^{j\phi^{(m)}[k]}] = 0$ that

$$E\left[Z^{(m)}[k]\right] = E\left[X^{(m)}[k]\right]\ E\left[e^{j\phi^{(m)}[k]}\right] = 0. \tag{4.23}$$

Since $Z^{(m)}[k]$ is zero-mean, the covariance between $Z^{(m)}[k]$ and $Z^{(n)}[l]$ is given by

$$\begin{aligned} E\left[Z^{(m)}[k]\left(Z^{(n)}[l]\right)^*\right] &= E\left[X^{(m)}[k]\left(X^{(n)}[l]\right)^* e^{j\phi^{(m)}[k]}e^{-j\phi_l^{(n)}}\right] &(4.24)\\ &= E\left[X^{(m)}[k]\left(X^{(n)}[l]\right)^*\right]\ E\left[e^{j\phi^{(m)}[k]}e^{-j\phi_l^{(n)}}\right]. &(4.25)\end{aligned}$$

Since $\{\phi^{(m)}[k]\}_{k=0}^{N-1}$ and $\{\phi_l^{(n)}\}_{l=0}^{N-1}$ are mutually independent for $m \neq n$ and $E[e^{j\phi^{(m)}[k]}] = 0$, expression (4.25) is zero when $m \neq n$.

This means that $Z^{(m)}[k]$ and $Z^{(n)}[k]$ are mutually uncorrelated for $m \neq n$. Because of (4.22), we infer that $z^{(m)}(t)$ and $z^{(n)}(t)$ are also uncorrelated. Since $z^{(m)}(t)$ and $z^{(n)}(t)$ are asymptotically jointly complex Gaussian distributed according to [9, ch. 4], the uncorrelatedness between $z^{(m)}(t)$ and $z^{(n)}(t)$ also implies their mutual independence.

Therefore, the CCDF of the PAR of the transmitted BSPTM-OFDM signal $z^{\bar{m}}(t)$ is obtained as

$$\begin{aligned} \Pr\left\{\mathrm{PAR}(z^{\bar{m}}(t)) > \gamma\right\} &= \Pr\left\{\min_{0\le m\le M-1}\{\mathrm{PAR}(z^m(t))\} > \gamma\right\}\\ &= \Pr\left\{\mathrm{PAR}(z^{(m)}(t)) > \gamma,\ \forall\ 0 \le m \le M-1\right\}\\ &= \left[\Pr\left\{\mathrm{PAR}(z^{(m)}(t)) > \gamma\right\}\right]^M\\ &= \left[1 - e^{-e^{-\gamma}N\sqrt{\frac{\tilde{\lambda}}{N}\log N}}\right]^M, &(4.26)\end{aligned}$$

where $\tilde{\lambda} = 4\pi^2/3 - \pi^2(1-\beta/P)^2$.

4.2.2.4 Blind Detection of $\theta_0^{(\bar{m})}$

At the receiver, we need to determine the optimum index $\bar{m}$. Let us replace the $X[k]$ in (4.9) by the $Z^{(\bar{m})}[k]$ of (4.17) and write:

$$\begin{aligned} Y[k] &= Z^{(\bar{m})}[k]H[k] + V[k]\\ &= \begin{cases} B[k]e^{j\phi_k^{(\bar{m})}}H[k] + V[k], & k \in \mathbf{\Omega}_{\bar{m}},\\ S[k]e^{j\phi_k^{(\bar{m})}}H[k] + V[k], & k \in \mathbf{\Omega}_{\bar{m}}^{\perp}. \end{cases} &(4.27)\end{aligned}$$

Our task here is to detect $\theta_0^{(\bar{m})}$ (or equivalently, $\bar{m}$) from $\{Y[k]\}_{k=0}^{N-1}$, knowing the candidate set of locations in $\mathbf{\Theta}$.

We will utilize the following assumptions in our discussions next:

1. $s[n]$, $v[n]$, and $h[n]$ are mutually independent,
2. $h[n]$ is i.i.d. zero-mean with variance σ_h^2,
3. $|B[k]|^2 = \sigma_p^2$ is constant $\forall k \in \mathbf{\Omega}_{\bar{m}}$ (equi-powered pilots).

From assumption 2, we infer that $H[k]$ has mean zero and variance $L\sigma_h^2$, $\forall\, k$. We also recall the following notations: $\sigma_s^2 = E[|S[k]|^2]$ and $\sigma_v^2 = E[|V[k]|^2]$. It follows from (4.27) that

$$E[|Y[k]|^2] = \begin{cases} \sigma_p^2\ (L\sigma_h^2) + \sigma_v^2, & k \in \mathbf{\Omega}_{\bar{m}}, \\ \sigma_s^2\ (L\sigma_h^2) + \sigma_v^2, & k \in \mathbf{\Omega}_{\bar{m}}^{\perp}. \end{cases} \tag{4.28}$$

Next, let us write $k = iR + r$, where $0 \le i \le P-1$ and $0 \le r \le R-1$, and denote by $Y_i[r] = Y[iR + r]$ the ith sub-record (of length-R) of $Y[k]$. It follows from (4.28) that

$$\rho[r] = \frac{1}{P} E\left[\sum_{i=0}^{P-1} |Y_i[r]|^2\right] = \begin{cases} \sigma_p^2\ (L\sigma_h^2) + \sigma_v^2, & r = \theta_0^{(\bar{m})}, \\ \sigma_s^2\ (L\sigma_h^2) + \sigma_v^2, & r \neq \theta_0^{(\bar{m})}. \end{cases} \tag{4.29}$$

Since $\sigma_p^2 > \sigma_s^2$, we infer from (4.29) that

$$\theta_0^{(\bar{m})} = \underset{r \in \mathbf{\Theta}}{\operatorname{argmax}} \{\rho[r]\}. \tag{4.30}$$

In practice, we estimate $\rho[r]$ as

$$\hat{\rho}[r] = \frac{1}{P} \sum_{i=0}^{P-1} |Y_i[r]|^2, \tag{4.31}$$

and find the optimum pilot location via

$$\hat{\theta}_0^{(\bar{m})} = \underset{r \in \mathbf{\Theta}}{\operatorname{argmax}} \{\hat{\rho}[r]\}. \tag{4.32}$$

Since the receiver has the knowledge of $\mathbf{\Theta}$, from $\hat{\theta}_0^{(\bar{m})}$, a simple lookup table search yields $\hat{\bar{m}}$.

Even if $|H[iR+r]|^2 \geq 0$ exhibits a deep null at $r = \theta_0^{(\bar{m})}$ for a particular sub-record i, since P sub-records are involved in the averaging in (4.31) and $\sigma_p^2 \gg \sigma_s^2$, $\hat{\rho}[r]$ is still likely to peak at $r = \theta_0^{(\bar{m})}$.

The side information $\bar{m}$ is critical for decoding at the receiver. If $\hat{\bar{m}}$ is inaccurate for a particular OFDM block, the BER will be high for that block. We point out however, that the finite alphabet nature of $\mathbf{\Theta}$ makes it less likely for errors to occur in $\hat{\theta}_0^{(\bar{m})}$. When the SNR is so low that σ_v^2 dominates the other terms on the RHS of (4.29), $\rho[r]$ at $r = \theta_0^{(\bar{m})}$ and at $r \neq \theta_0^{(\bar{m})}$ become less distinguishable, and hence accurate detection of $\bar{m}$ becomes difficult. As we will see in the simulations section, at medium to high SNR levels, the detector in (4.32) performs quite reliably, especially when σ_p^2/σ_s^2 is high.

4.2.3 Simulations

In the examples in this section (except when specified otherwise), we assume that the number of sub-carriers $N = 128$, the length of the FIR channel $L = 4$, the number of pilot tones $P = L = 4$, and the power allocation factor $\beta = 0.15$ (c.f. (4.13)). Except for Section 4.2.3.3, the phase table consists of i.i.d. $\{0,\ \pi\}$ entries with equal probability; in other words, we have an $\{e^{j\phi^{(m)}[k]}\}$ table consisting of i.i.d. $\{1,\ -1\}$ entries with equal probability. Such a sign change table is pre-determined and is stored at both the transmitter and the receiver. The $N-P$ information sub-symbols were independently drawn from a QPSK constellation with Gray coding. Under the unit channel energy constraint $\sum_{n=0}^{L-1} E[|h[n]|^2] = 1$, the signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{\mathcal{P}_{dc}}{\sigma_v^2}, \tag{4.33}$$

where $\mathcal{P}_{dc}$ is the total amount of DC power consumed by the PA and σ_v^2 is the variance of the additive white Gaussian noise. The effective SNR, which directly affects the BER, can be expressed as $\text{SNRe} = \mathcal{P}_t/\sigma_v^2$, where $\mathcal{P}_t$ is the average output power of the PA with the input signal $z^{(\bar{m})}(t)$. If an ideal linear PA is used and the signal is to be amplified undistorted, $\mathcal{P}_t$ is proportional to $\mathcal{P}_{dc}/\text{PAR}$ if linear scaling is employed [47]. Therefore, $\text{SNRe} \propto \mathcal{P}_{dc}/\text{PAR}$, and the benefit of PAR reduction is realized as an increase in SNRe.

4.2.3.1 PAR Reduction Performance

In this example, we approximate the continuous-time PAR of (2.3) by evaluating the discrete-time PAR of the 4-times oversampled OFDM signal [62]. 10^6 independent Monte Carlo trials were conducted.

Fig. 4.7 shows the empirical CCDF curves (solid lines) of the PAR of the transmitted signal $z^{(\bar{m})}(t)$ for different number of selections, M, along with the theoretical CCDFs (dash-dotted lines) obtained from (4.26). $M = 1$ corresponds to the original PTAM-OFDM case. Fig. 4.7 shows that the empirical and the theoretical CCDFs are quite close. We observe that when $M = 8$, the proposed algorithm could achieve 3.5 dB of PAR reduction (as compared with the $M = 1$ case) at the CCDF level of 10^{-4}. We also see from Fig. 4.7 that the larger the M, the smaller the resulting PAR. On the other hand, the computational complexity increases as M increases. There is also a diminishing return in the PAR reduction capability as M further increases. As a rule of thumb, we recommend to use $\min\{R, 4\} \leq M \leq \min\{R,\ 8\}$.

4.2.3.2 Blind Detection of $\bar{m}$

We provide an example here to illustrate the blind detection of $\bar{m}$ from $|Y[k]|^2$. In this example, SNR = 0 dB, $R = 32$, $\mathbf{\Theta} = \{0,\ 4,\ 8,\ 12,\ 16,\ 20,\ 24,\ 28\}$, and thus $M = 8$. The channel taps are assumed to be i.i.d. complex Gaussian distributed with zero-mean and variance $\sigma_h^2 = 1/L$ (i.e., Rayleigh fading), and $L = 4$. Fig. 4.8 shows $|H[k]|$ vs. k for one realization of the Rayleigh fading channel with time-domain coefficients

$$\mathbf{h} = [0.2774 - 0.4545j, -0.4988 + 0.1837j, 0.1189 + 0.1105j, -0.0751 - 0.6340j]^T, \quad (4.34)$$

which exhibits several deep nulls in the frequency domain. Table 4.2 shows the PAR$\{z^{(m)}(t)\}$ values for one particular OFDM block, with $0 \leq m \leq 7$. We observe that $m = 6$ corresponds to the lowest PAR value, thus the optimum pilot tone location parameter was $\theta_0^{(6)} = 24$. At the receiver, we first calculate $|Y[k]|^2$. In Fig. 4.9, for each sub-record $|Y_i[r]|^2$, we use circles to indicate the values corresponding to the M candidate locations $r \in \mathbf{\Theta}$. From the $\hat{\rho}[r]$ plot, we found $\hat{\theta}_0^{(\bar{m})} = 24$ (or equivalently, $\hat{\bar{m}} = 6$), which was indeed the true $\theta_0^{(\bar{m})}$ that was used during transmission.

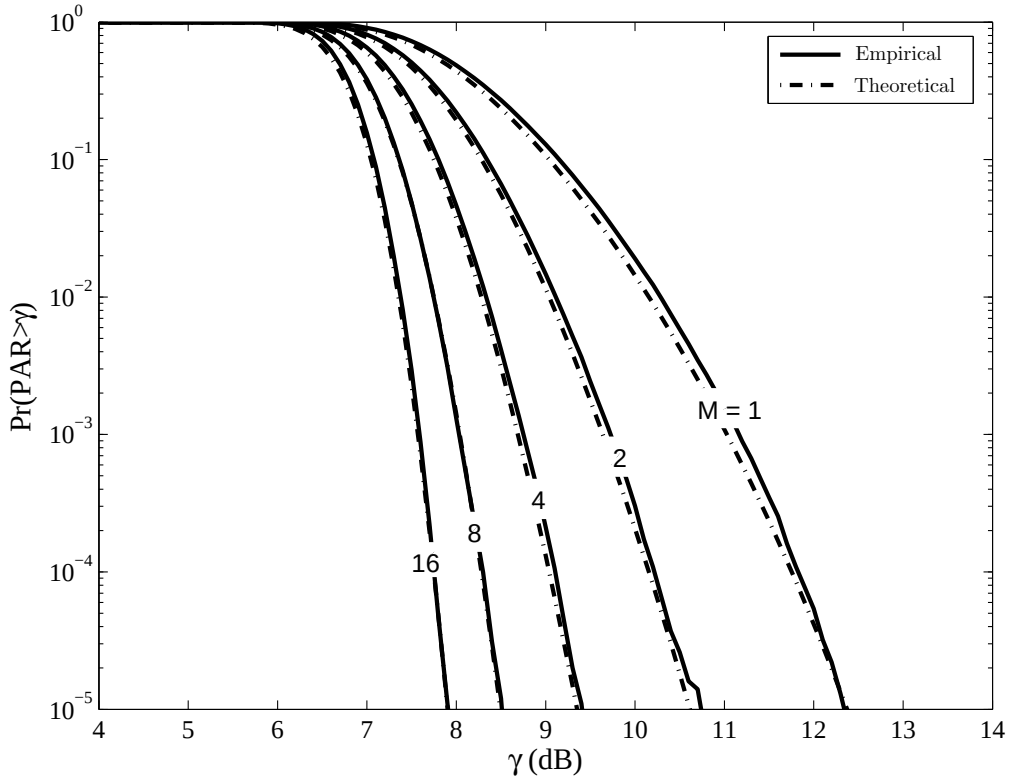


Figure 4.7: CCDF of the PAR of the BSPTM-OFDM signal for varying values of M with $N = 128$, $P = 4$, and $\beta = 0.15$.

Table 4.2: PAR$\{z^{(m)}(t)\}$ for one OFDM symbol, $0 \leq m \leq 7$.

m	0	1	2	3	4	5	6	7
PAR (dB)	7.4424	7.3252	10.3441	8.1139	8.5495	8.0270	*6.6928*	7.3648

To illustrate the performance on the blind detection of $\bar{m}$, we list in Table 4.3 the detection error rates for varying values of β, M, and SNR, calculated by averaging over 10^5 Monte Carlo trials. It is evident from Table 4.3 that the larger the β, the smaller the error rate. This is because when β is larger, $\rho[r]$ of (4.29) stands out better at $r = \theta_0^{(\bar{m})}$. Moreover, we observe from Table 4.3 that the larger the M, the higher the error rate in detecting $\theta_0^{(\bar{m})}$. This is because there are more competing candidate m's when M is larger. When β is not too small (e.g., $\beta > 0.1$), the error rate can be quite small for SNR > 0 dB. However, β cannot be too large either, since when too much power is devoted to the pilot tones instead of the information sub-symbols, the receiver becomes vulnerable to the channel distortions

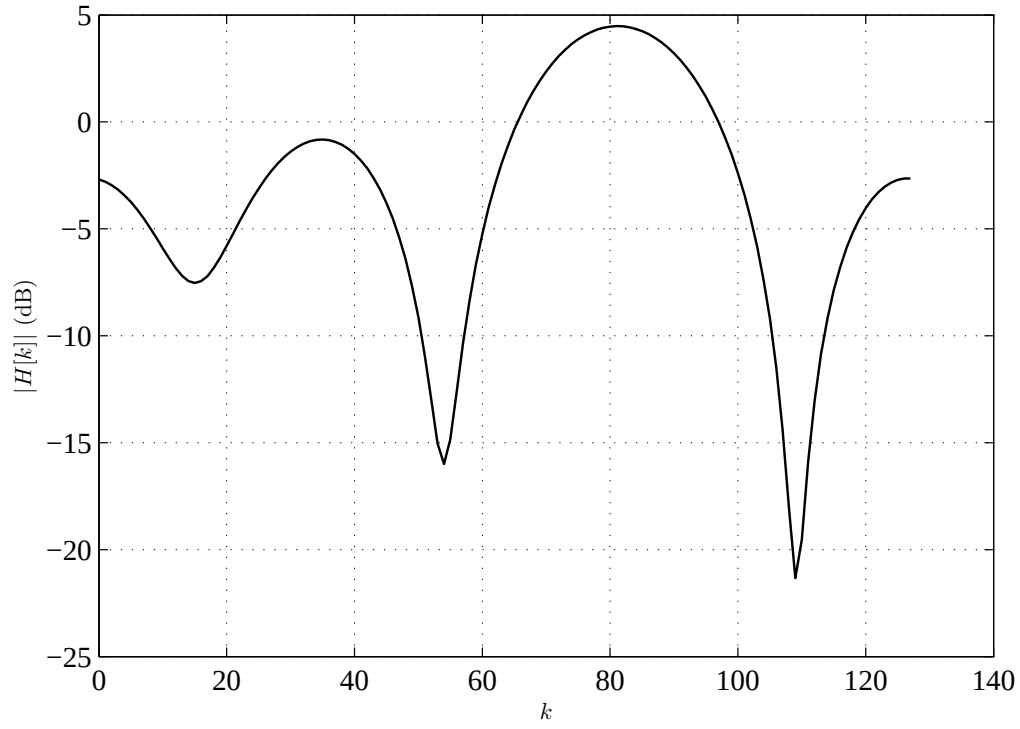


Figure 4.8: Magnitude response of one realization of the Rayleigh channel.

and the additive noise.

4.2.3.3 Comparison with Ref. [28] on the detection of $\bar{m}$

In this example, we compare the performance of the proposed BSPTM method with that of [28] in the presence of Rayleigh fading. The simulation parameters were the same as in the previous example, except that the phases $\{\phi^{(m)}[k]\}$ were i.i.d. uniformly distributed in $[-\pi,\ \pi)$ (the method of Ref. [28] does not work when the phases have a discrete distribution). Note that the ML decoder of [28] needs the CSI in order to detect the optimum phase sequence index $\bar{m}$, but BSPTM does not. Table 4.4 compares the error rates in detecting $\bar{m}$ between the method of Ref. [28] and our proposed BSPTM technique. Please note that we had assumed perfect CSI for the method of [28] but no CSI for BSPTM. Despite the favorable setup for [28], BSPTM is clearly more robust. Moreover, the ML decoder of [28] has a higher computational complexity than BSPTM. For example, if the

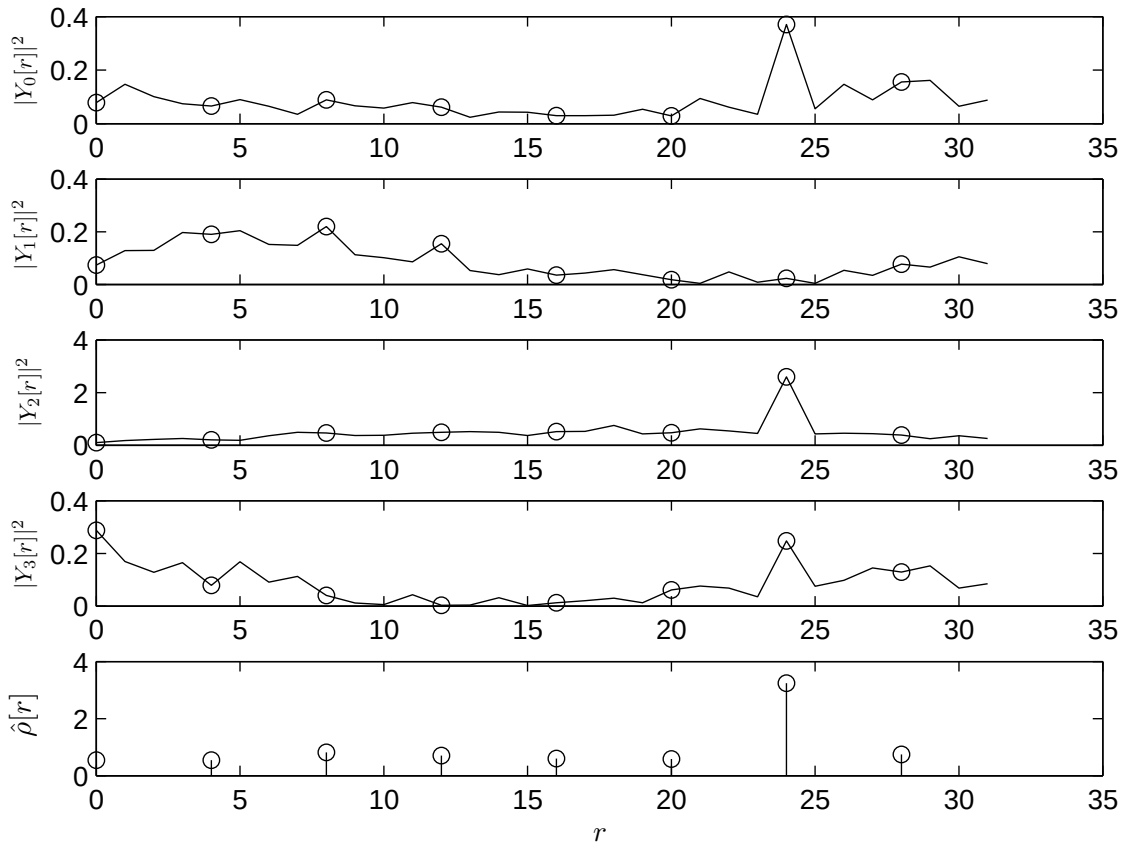


Figure 4.9: From top to bottom: $|Y_0(r)|^2$, $|Y_1(r)|^2$, $|Y_2(r)|^2$, $|Y_3(r)|^2$, and $\hat{\rho}[r]$ with $r \in \boldsymbol{\theta} = \{0,\ 4,\ 8,\ 12,\ 16,\ 20,\ 24,\ 28\}$ over $r \in \boldsymbol{\Theta}$. $\hat{\rho}[r]$ peaked at $r = 24$, agreeing with the true $\theta_0^{(6)} = 24$ ($\bar{m} = 6$) used in the transmission.

frequency-domain OFDM sub-symbols are drawn from a 16QAM constellation, the ML decoder requires $16MN$ magnitude-squared ($|\cdot|^2$) operations, whereas BSPTM only needs N of them.

4.2.3.4 BER Performance

Next, we compare the BER performance of BSPTM-OFDM with that of PTAM-OFDM for $N = 128$, $P = 4$, $\beta = 0.15$, and $M = 8$. The receiver uses a zero-forcing equalizer and a suboptimal but simple hard-decision decoder [50]. Similar to [50], we tried two types of channels: a fixed FIR channel with tap coefficients in (4.34) and a Rayleigh fading channel with i.i.d. complex Gaussian taps. The BER was evaluated by averaging over 10^5 Monte Carlo trials.

Table 4.3: Error rate in detecting $\bar{m}$ for varying β, M, and SNR $= \mathcal{P}_{dc}/\sigma_v^2$ values.

SNR		0 dB	5 dB	10 dB	20 dB
$M = 4$	$\beta = 0.1$	6.05%	0.32%	0%	0%
	$\beta = 0.2$	0.39%	0%	0%	0%
	$\beta = 0.3$	0.06%	0%	0%	0%
$M = 8$	$\beta = 0.1$	8.71%	0.67%	0.01%	0%
	$\beta = 0.2$	0.46%	0%	0%	0%
	$\beta = 0.3$	0.08%	0%	0%	0%
$M = 16$	$\beta = 0.1$	12.37%	0.77%	0.02%	0%
	$\beta = 0.2$	0.59%	0%	0%	0%
	$\beta = 0.3$	0.11%	0%	0%	0%

Table 4.4: Error rates in detecting $\bar{m}$ when $N = 128,\ P = 4,\ M = 8,\ \beta = 0.15$.

SNR	0 dB	5 dB	10 dB	20 dB
Ref. [28]	35.28%	5.02%	0.46%	0.02%
BSPTM	1.42%	0.09%	0%	0%

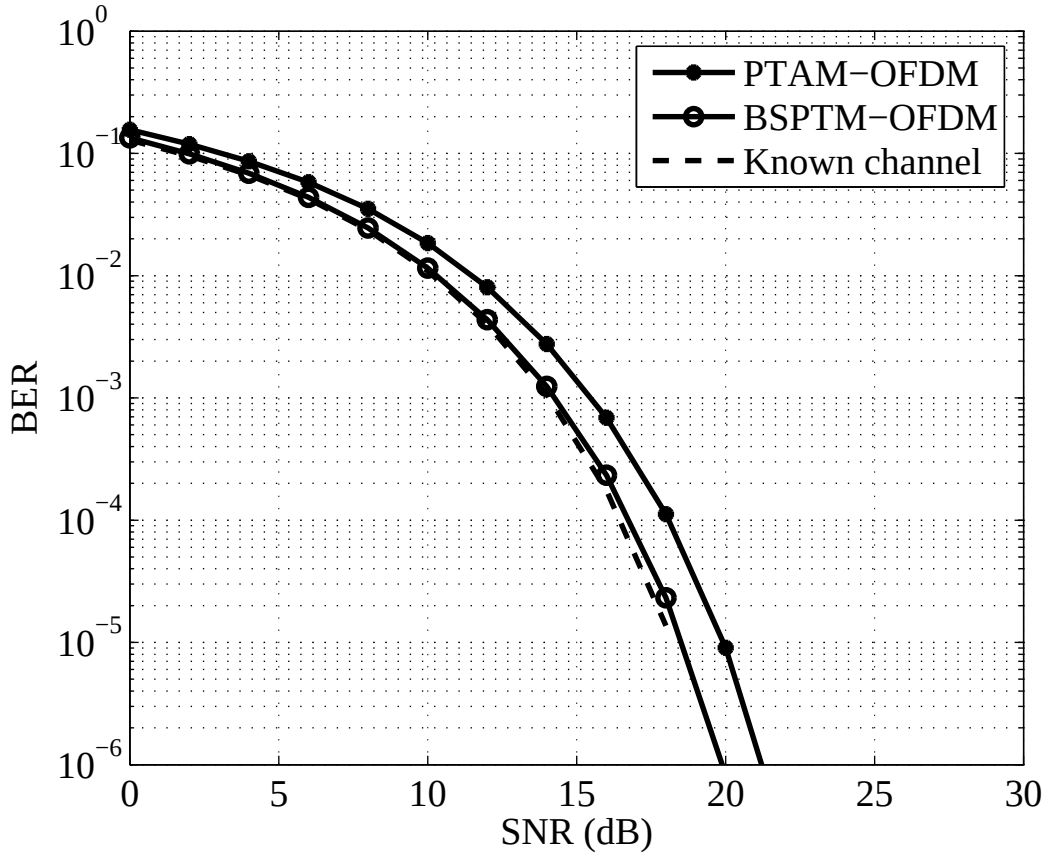


Figure 4.10: BER of PTAM-OFDM and BSPTM-OFDM for the fixed channel with $N = 128$, $P = L = 4$, $M = 8$, and $\beta = 0.15$.

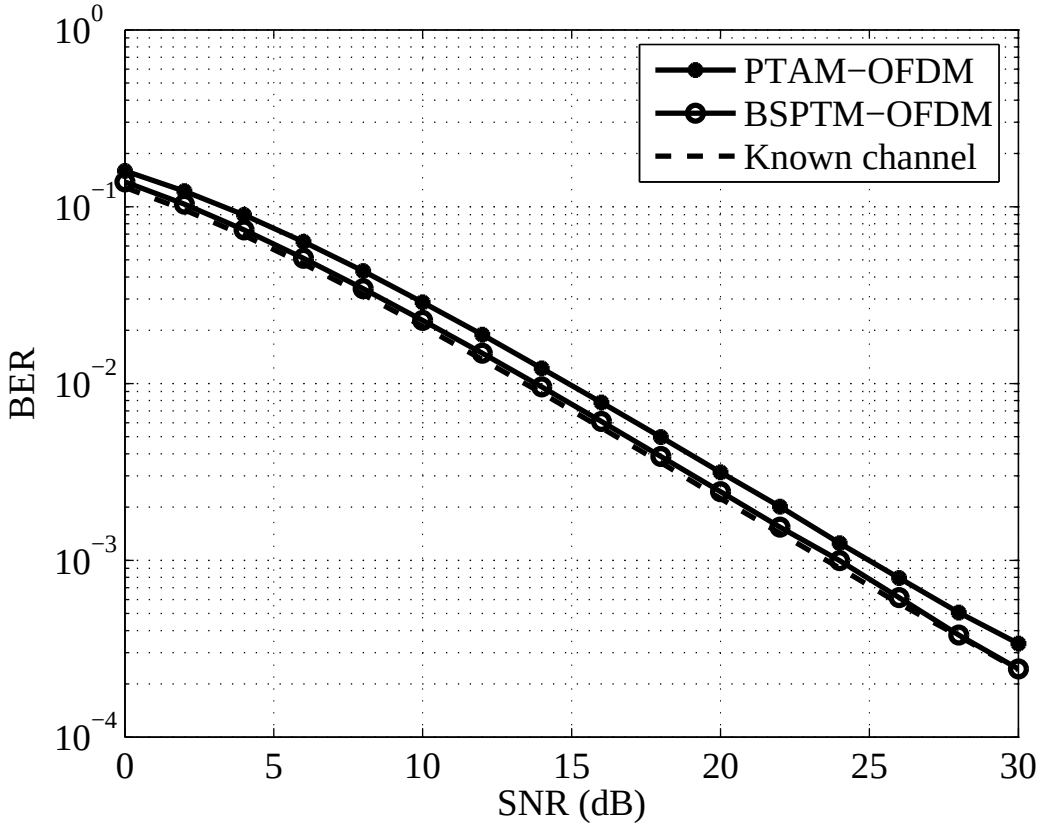


Figure 4.11: BER of PTAM-OFDM and BSPTM-OFDM for the Rayleigh channel with $N = 128$, $P = L = 4$, and $M = 8$.

Fig. 4.10 shows the BER performance of the proposed BSPTM technique and that of PTAM-OFDM for the fixed channel case. Fig. 4.11 shows a similar comparison for the Rayleigh fading case. We can see from both figures that the PTAM-OFDM performance is only $1 - 2$ dB away from the known channel case, which can serve as a benchmark. However, our proposed BSPTM-OFDM method offers even better BER performance, which approaches the performance of the known channel case for both the fixed and the Rayleigh fading channels. Such superior performance is possible, since we have taken advantage of the reduction in the PAR to boost the average transmission power for the same amount of DC power. Specifically, we have kept the peak power of an OFDM block fixed, but adjusted the average power according to the actual PAR value of the block. This linear scaling approach [47] ensures the most efficient utilization of the PA; in other words, the average transmit power is made proportional to $\mathcal{P}_{dc}$/PAR [47]. Eventually, the benefit of

PAR reduction is realized as decrease in the BER.

4.3 Blind SLM for Forward Link CDMA

In this section, we propose two novel SLM techniques[2] that are unique to the forward link CDMA signal. Both methods are distortionless and are blind in the sense that the side information used by the transmitter to achieve the PAR reduction is not transmitted in order to preserve the information rate. We describe methods to detect the side information from the received data, prior to symbol recovery.

4.3.1 SLM for Forward Link CDMA

The complex envelope of the nth CDMA forward link symbol is defined as [35, 36]

$$x(t) = \sum_{l=0}^{L-1} \sum_{k=1}^{K} A_k S_n[k] W_k[l] (P_I[l] + jP_Q[l]) h_T(t - lT_c), \ (n-1)T_s \le t \le nT_s, \qquad (4.35)$$

where

A_k: amplification factor for the kth channel;

$S_n[k]$: the kth channel data in the nth symbol period;

$W_k[l]$: the lth chip of the Walsh code assigned to the kth channel;

$P_I[l]$: the lth chip of the I-phase short PN sequence;

$P_Q[l]$: the lth chip of the Q-phase short PN sequence;

$h_T(t)$: impulse response of the transmit filter;

T_c: chip duration;

T_s: symbol duration;

L: number of chips per symbol, i.e., T_s/T_c;

K: number of active channels.

Eq. (4.35) provides a general baseband model for the forward link CDMA system. $S_n[k]$ is BPSK for IS-95 and QPSK for CDMA2000 [35, 36].

IAR of the CDMA signal is defined as

$$\text{IAR} = \frac{\mathcal{P}(t)}{\mathcal{P}_{av}}, \qquad (4.36)$$

[2]The blind SLM techniques described in this section are protected by the U.S. Utility Patent Application: "Blind Selected Mapping Techniques for Crest Factor Reduction of Forward Link CDMA Signals," by N. Chen and G. Tong Zhou, filed on December 12, 2005 with Serial No. 11/301,381.

where $\mathcal{P}(t) = |x(t)|^2$ and $\mathcal{P}_{av} = E[|x(t)|^2]$.

In this work, we define PAR as the value γ corresponding to

$$\Pr(\text{IAR} > \gamma) = 10^{-4}. \tag{4.37}$$

The objective here is to reduce the γ defined above.

SLM was first proposed as a simple and effective method to reduce the PAR of OFDM signals. In SLM, M equivalent representations of the same signal $x(t)$, $\{x^{(m)}(t)\}_{m=1}^{M}$, are first obtained. Assume that the average power is invariant; i.e., $E[|x^{(m)}(t)|^2] = E[|x(t)|^2]$, $\forall\ m$, we only need to pay attention to the peak power

$$p^{(m)} = \max_t |x^{(m)}(t)|^2, \tag{4.38}$$

where the maximization is over one OFDM block. Denote

$$\bar{m} = \underset{1 \le m \le M}{\arg\min}\ \ p^{(m)}. \tag{4.39}$$

SLM pertains to the transmission of $x^{(\bar{m})}(t)$ in place of $x(t)$. Next, we extend the SLM framework to reduce the PAR of the forward link CDMA signals.

4.3.2 Blind Phase SLM

In (4.35), let us replace $S_n[k]$ by

$$S_n^{(m)}[k] = S_n[k]\ e^{j\phi^{(m)}[k]}, \tag{4.40}$$

where the table of possible $e^{j\phi^{(m)}[k]}$ values is available at both the transmitter and the receiver. We then obtain equivalent signal representations $x^{(m)}(t)$ whose peak power $p^{(m)}$ can be smaller or larger than that of $x(t)$. Since the PAR of $x^{(\bar{m})}(t)$ will never exceed that of $x(t)$, we thus have a distortionless PAR reduction method.

4.3.2.1 Phase Table Design

The phases $\phi^{(m)}[k]$ used in Eq. (4.40) come from a pre-determined table comprising M rows and K columns (we assume without loss of generality here that each $S_n[k]$ is arranged as a row vector in k). Each row corresponds to a different $x^{(m)}(t)$. The table may be formed by

Table 4.5: An example of the 8 by 4 "random" phase table in phase SLM for CDMA.

	$k=1$	$k=2$	$k=3$	$k=4$
$m=1$	0	0	0	0
$m=2$	0	$-\pi/2$	$-\pi$	$\pi/2$
$m=3$	$\pi/2$	$-\pi$	0	$-\pi/2$
$m=4$	0	$-\pi/2$	$\pi/2$	$-\pi$
$m=5$	$-\pi/2$	$\pi/2$	0	$-\pi$
$m=6$	$\pi/2$	$-\pi/2$	0	$-\pi$
$m=7$	$-\pi/2$	0	$-\pi$	$\pi/2$
$m=8$	0	$-\pi$	$-\pi/2$	$\pi/2$

randomly selecting phases from $[-\pi, \pi)$ (or equivalently, $[0, 2\pi)$ etc, since $e^{j(\phi+2\pi)} = e^{j\phi}$). In simulations, a phase table formed in a different manner was used. First, a K-element vector $\mathbf{a} = \pi \times [-1, \ -1+2/K, \ \cdots, \ 1-2/K]$ is formed so the elements are taken from $[-\pi, \pi)$ with step size $2\pi/K$. There are $K!$ different permutations of the elements of $\mathbf{a}$. Let the mth row of the phase table be formed from the mth permutation of $\mathbf{a}$. Since generally $M < K!$, the rows of the phase table do not repeat. The first row has all zero phases, corresponding to the original signal without any phase rotations. In practice, $e^{j\phi^{(m)}[k]}$ are pre-calculated and stored in the table. Since PAR reduction adds to the processing overhead, it may be desirable to carry out PAR reduction for every block of N symbols. In that case, the above phase table can be re-used for every symbol block. Alternatively, an $M \times (NK)$ phase table may also be created in which the NK columns do not repeat.

Example: The following is an example of how to construct the "random" phase table with $K = 4$ and $M = 8$, and apply it to the phase selected mapping algorithm. First, a 4 element row vector $\mathbf{a} = [-\pi, -\pi/2, 0, \pi/2]$ is formed. There are $4 \times 3 \times 2 = 24$ different permutations of the elements of $\mathbf{a}$. The first row of the "random" phase table is set as zeros, corresponding to the original signal without any phase rotations. Then, randomly select 7 permutations of $\mathbf{a}$ to fill row 2 to row 8 in Table 4.5. The nth multichannel symbol has $K = 4$ elements, which are phase rotated by each row of Table 4.5 to yield $M = 8$ equivalent signal representations as shown in Fig. 4.12. When the SLM frame size $N = 10$, measure the peak in each branch of Fig. 4.12 over a 10 symbol period, and record the peak-to-average power ratio in Table 4.6. Therefore, the output signal of the third branch is transmitted

Table 4.6: An example of the measured PARs in phase SLM for CDMA.

	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$
$\text{PAR}(x^{(m)}(t))$	5.1517	4.9218	3.6623	4.9608	5.1829	4.1500	5.3899	4.2546

and $\bar{m} = 3$.

The receiver must know the side information $\bar{m}$ in order to decode. There are various ways to manage the side information issue.

One method is to transmit the side information about $\bar{m}$ in an overhead channel. Since the information about $\bar{m}$ is of critical importance to the receiver, it should be protected by channel coding. If $M = 16$, the side information $\bar{m}$ occupies 4 bits in the uncoded case. Thus, the amount of overhead associated with the transmission of the side information can be significant.

Presented below are novel methods to blindly detect $\bar{m}$ at the receiver, by relying on orthogonal or near orthogonal properties of the CDMA channelization codes.

4.3.2.2 Blind Detection Using Pilot Channel

Since the pilot channel has $k = 1$ and $S_n[k] = S_n[1] = 1$, $\forall n$, $W_k[l] = W_1[l] = 1$, $\forall l$, the received pilot channel signal corresponds to $(P_I[l] + jP_Q[l])e^{j\phi^{(\bar{m})}[1]}$; the receiver sees that every chip of the complex PN sequence is rotated by the same angle $\phi^{(\bar{m})}[1]$. Using the correlation properties of the PN sequence, it is thus possible to determine the optimum phase sequence index $\bar{m}$ used in the transmission, since

$$\sum_{l=1}^{L}(P_I[l] + jP_Q[l])e^{j\phi^{(\bar{m})}[1]}(P_I[l] - jP_Q[l]) = 2L\ e^{j\phi^{(\bar{m})}[1]}. \tag{4.41}$$

We require that the phase table has $e^{j\phi^{(m)}[1]} \neq e^{j\phi^{(i)}[1]}$ for $m \neq i$ in order to ensure the unique identification of $\bar{m}$.

4.3.2.3 Blind Detection Using PN Offsets

Recall that in the CDMA forward link, each base station or sector identifies itself by employing one of $2^{15} = 32,768$ PN offsets. In order to avoid possible confusion in the PN offset caused by propagation delays from different base stations to a mobile receiver, the IS-95 standard specifies that the minimum separation between two PN offsets is 64 chips.

Consequently, $32,768/64 = 512$ PN offsets are available. Suppose that PN offset $64i$ is used to identify the ith base station, $i = 0, \ldots, 511$. We propose to associate a "cluster" of PN offsets $\{64i+m-1\}_{m=1}^{M}$ with base station i (see Fig. 4.13). If within the same geographical area, no more than 256 base stations are needed (the same set of PN offsets can be reused in a different region), we can plan the PN offsets such that $\{128i+m-1\}_{m=1}^{M}$ identify the ith base station. This way, as long as $M \leq 64$, the minimum separation between the PN offsets of different base stations is still at least 64 chips. With such PN offset planning, we can embed the phase sequence index m in the choice of the short PN sequences. In essence, we modify (4.35) as follows (see also Fig. 4.12):

$$x^{(m)}(t) = \sum_{l=0}^{L-1} \sum_{k=1}^{K} A_k \; S_n^{(m)}[k] \; W_k[l] \; (P_I^{(m)}[l] + jP_Q^{(m)}[l]) \; h_T(t - lT_c), \tag{4.42}$$

where $S_n^{(m)}[k]$ is given by (4.40) and the same index m is carried by the in-phase and quadrature PN sequences as well. We refer to this as the blind selected phase and PN offset mapping (BSPNM) technique. Since at the receiver, the PN sequence offset can be determined easily using cross-correlation techniques, blind detection of $\bar{m}$ using the same technique is straightforward. As compared to the method described in Section 4.3.2.2, this technique may have better performance in the low SNR and high co-channel interference scenarios since we can combine the traffic as well as the overhead channel data to generate a more reliable $\bar{m}$ estimate.

In order to avoid PN sequence aliasing, the following two conditions must be satisfied according to [75]:

$$d_s > 122\mathrm{W_{in}} + 2r, \tag{4.43}$$

where d_s is the distance between two base stations that use the same PN offset, $\mathrm{W_{in}}$ is the size of the PN offset search window in chips (i.e., system parameter SRCH_WIN_A), and r is the coverage radius of the home base station in meters. In addition, we must ensure

$$d_a < 244I - 122\mathrm{W_{in}} + 2r, \tag{4.44}$$

where d_a is the maximum distance between two base stations that use adjacent PN offsets, and I is the separation between two adjacent PN offsets in chips.

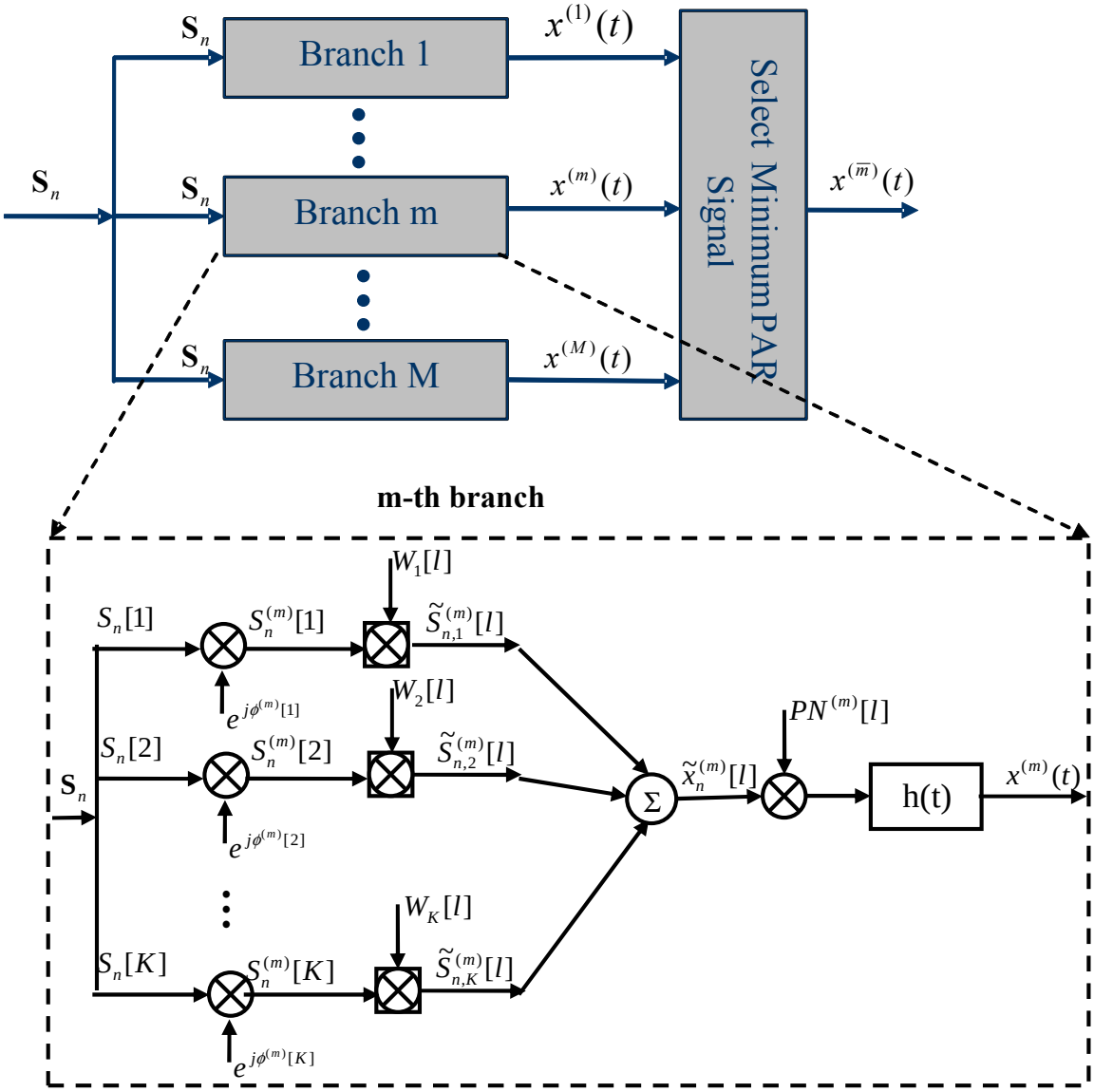


Figure 4.12: The mth branch of the baseband equivalent structure for BSPNM-CDMA transmission.

These conditions can be easily satisfied. For example, when $\mathrm{W_{in}} = 20$ chips, $r = 2,000$ meters, and $I = 64$ chips, we have

$$d_s > 6,440 \text{ meters}, \qquad d_a < 17,176 \text{ meters}.$$

Note that in our proposed method, the separation between adjacent PN offsets is reduced from I to $I - M - 1$ chips. Because of this, condition (4.44) changes to

$$d_a < 244(I - M - 1) - 122\mathrm{W_{in}} + 2r. \tag{4.45}$$

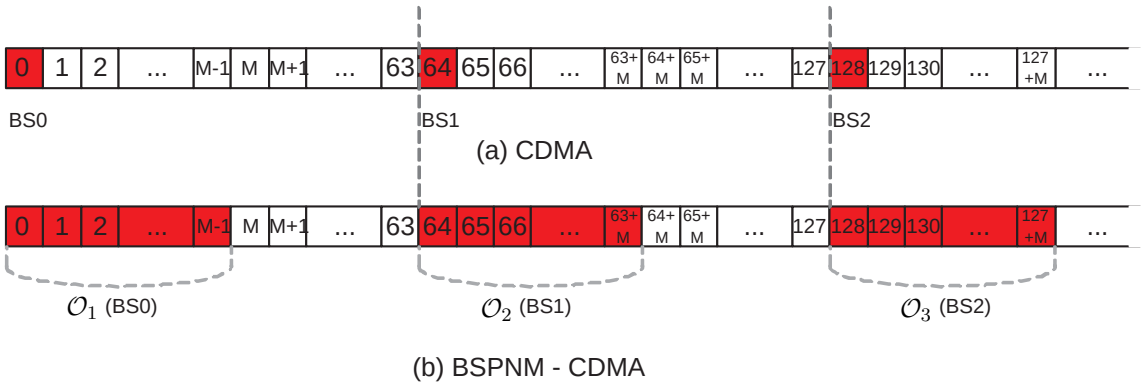


Figure 4.13: PN offset planning for the CDMA and the BSPNM-CDMA systems.

For the same example with $\mathrm{W_{in}} = 20$ chips, $r = 2,000$ meters, $I = 64$ chips, and $M = 16$, we now have

$$d_s > 6,440 \text{ meters}, \qquad d_a < 13,272 \text{ meters}.$$

The modified requirement on d_a can still be easily satisfied, as long as we assign adjacent PN offsets to those cells that are physically close and similar in size [4].

4.3.3 Blind Walsh Code SLM

It is well known (e.g., [7], [33]) that the PAR of a CDMA signal is a function of the Walsh code(s) used. Different from [35] and [58] which aim at finding one good set of Walsh codes, we propose to extend the SLM idea to Walsh code selection. Suppose that the system is not at full capacity and each traffic channel can use one of two alternative Walsh codes. Table 4.7 shows an example Walsh code assignment scheme; the same table is available at the transmitter and at the receiver. We remark that there are other ways to assign the Walsh codes; for example, we may assign $\{W_8, W_9\}$ to traffic channel #1, $\{W_{10}, W_{11}\}$ to traffic channel #2, etc. For 6 traffic channels, there are $2^6 = 64$ different combinations of the Walsh codes, each leading to a different PAR for the resulting forward channel signal. Fig. 4.14 shows the transmitter structure for the proposed blind Walsh code SLM technique. As we can see, for K channels with 2 Walsh code per channel, there are $M = 2^K$ different combinations of the Walsh codes, resulting in $M = 2^K$ forward channel signals. We then select $x^{(\bar{m})}(t)$ that has the lowest PAR to transmit (see Fig. 4.14). However, we may only use a small number $M < 2^K$ of the 2^K available combinations. These M combinations can

Table 4.7: A Walsh code allocation scheme.

Traffic Channel	Walsh Code Choices
1	W_8 or W_{33}
2	W_9 or W_{34}
3	W_{10} or W_{35}
4	W_{11} or W_{36}
5	W_{12} or W_{37}
6	W_{13} or W_{38}
$\vdots$	$\vdots$

be selected empirically to have good PAR properties [35, 58]. We then select $x^{(\bar{m})}(t)$ that has the lowest PAR among $\{x^{(m)}(t)\}_{m=1}^{M}$, to transmit.

Thanks to the orthogonality property of the Walsh codes, the mobile receiver will be able to figure out which one of the two Walsh codes was used for its traffic channel, by cross-correlating either Walsh code with the received data.

4.3.4 Simulations

In this section, we demonstrate the performance of the proposed PAR reduction techniques. In the simulations, we approximated the continuous-time IAR of (4.36) by evaluating the discrete-time IAR of the 4-time oversampled (over the chip-rate) CDMA signal. In each example, the CCDF was obtained from 10^4 CDMA symbol periods. The PAR was measured from the IAR as $\Pr\{\text{IAR} > \text{PAR}\} = 10^{-4}$. We used the channel parameters given by Table 6.5.2-1 of 3GPP2 C.S0010-B v2.0 "Base Station Test Model, Nominal for Main Path" (3 overhead channels plus 6 traffic channels).

4.3.4.1 Phase SLM

Fig. 4.15 shows the CCDF of the IAR for the IS-95 signal, before and after BSPNM. The performance of the technique proposed by Lee and Miller in [36] is also shown for comparison. We observe that the larger the M (more mapping selections), the better the PAR reducing capability of SLM, but there is a diminishing return in increasing M beyond say, $M = 32$. With $M = 32$, the PAR was reduced by more than 4.4 dB at the CCDF level of 10^{-4}.

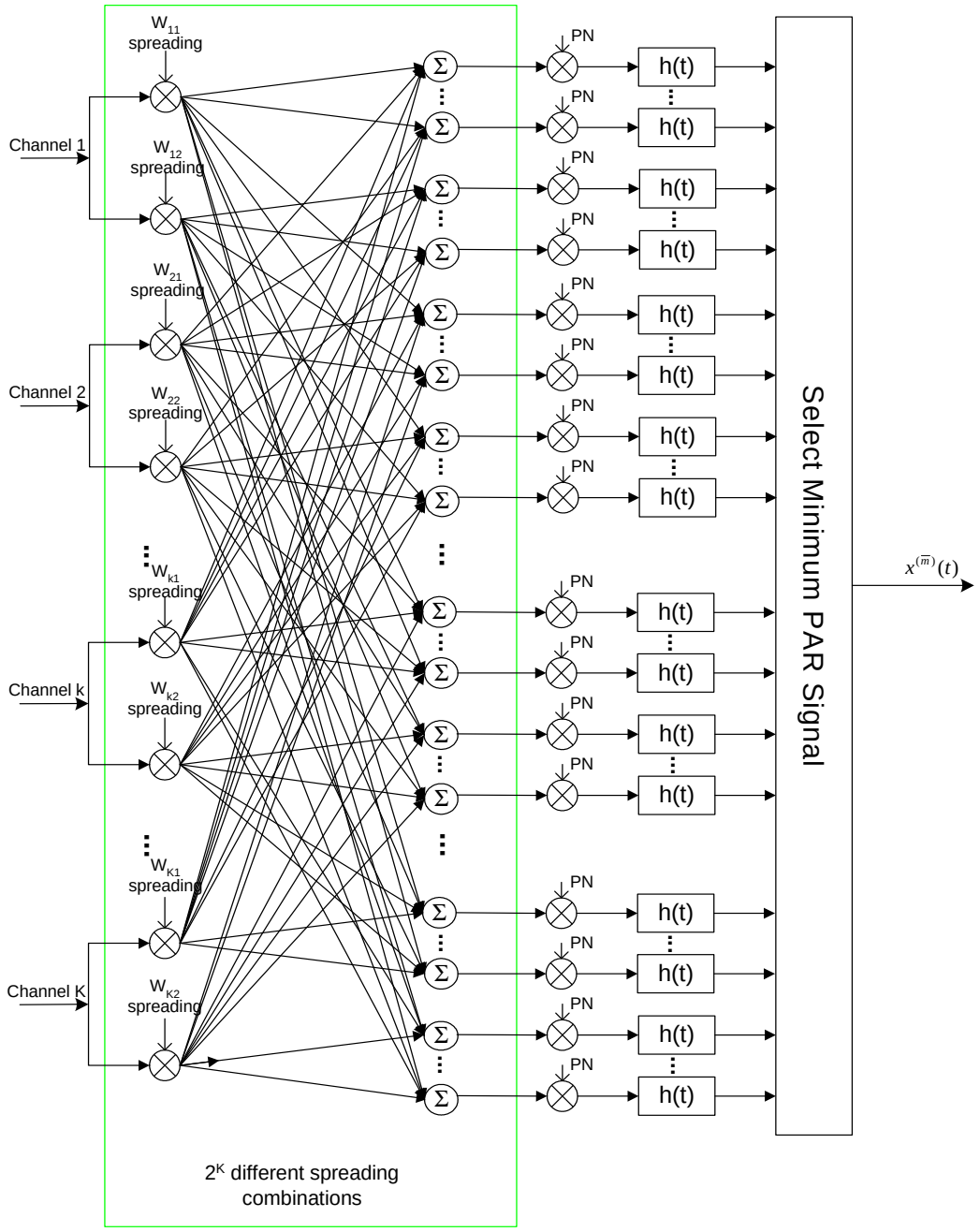


Figure 4.14: The proposed forward link CDMA transmitter with Walsh code SLM.

4.3.4.2 Walsh Code SLM

Fig. 4.16 shows the CCDF of the IAR for the CDMA2000 signal, before and after the Walsh code SLM algorithm. The code allocation scheme for the 6 traffic channels is shown in Table 4.7. We tested with the following parameters $M = 2$ and $M = 16$. Performance of the technique proposed in [36] is also shown for comparison. We see that the larger the

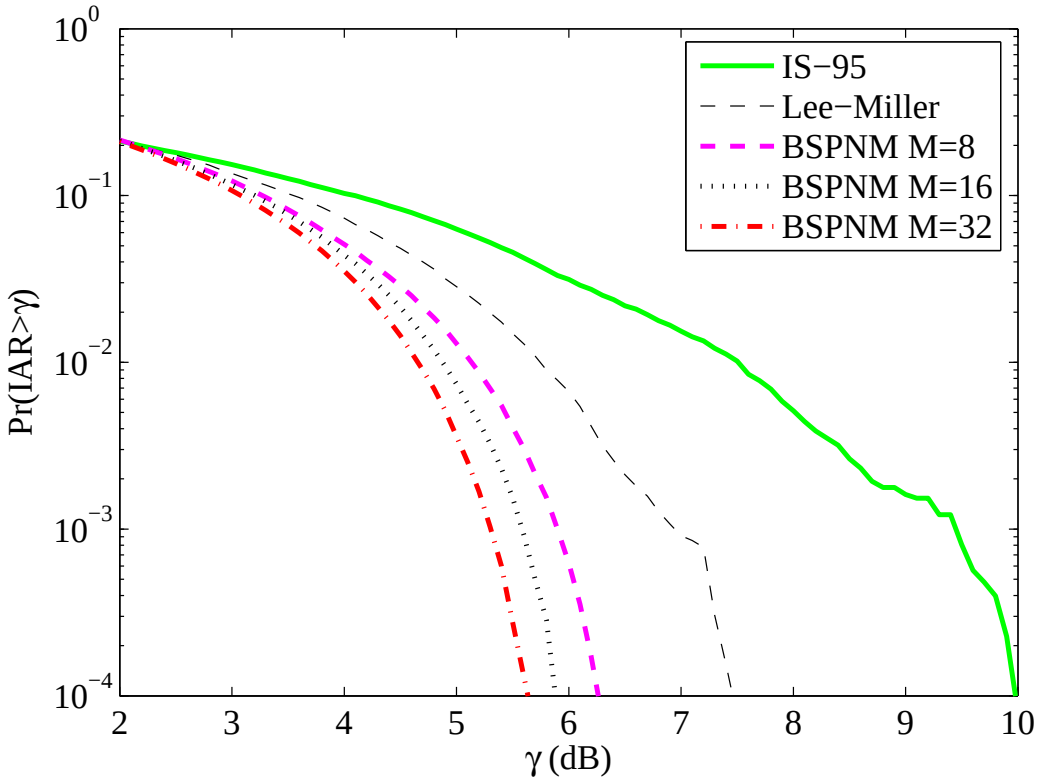


Figure 4.15: PAR reduction performance of BSPNM for IS-95.

M, the greater the PAR reducing capability.

From the simulations, we see that phase SLM is more powerful than Walsh code SLM in reducing the crest factors, but the latter method requires less modification to the network protocols. For both methods, the PAR is reduced more (statistically) if M is larger, but the computational complexity also increases with increasing M.

4.3.4.3 PAR versus BSPNM Frame Size N

In this example, we fix the number of selections ($M = 8$) and the number of active channels ($K = 14$), and examine the effect of the frame size N on the PAR reduction capability. Fig. 4.17 shows the CCDF of the IAR of the BSPNM-CDMA2000 signal with varying values of N. As illustrated in this figure, when $N = 1$, the PAR can be reduced by 2.9 dB. However, as N increases, the PAR reduction capability decreases. For a satisfactory PAR reduction performance, we would like N to be small. On the other hand, as we discussed in Section 4.3.2, a smaller N requires more frequent detection on the selected index $\bar{m}$, incurring a relatively higher complexity at the mobile receiver. Therefore, there is

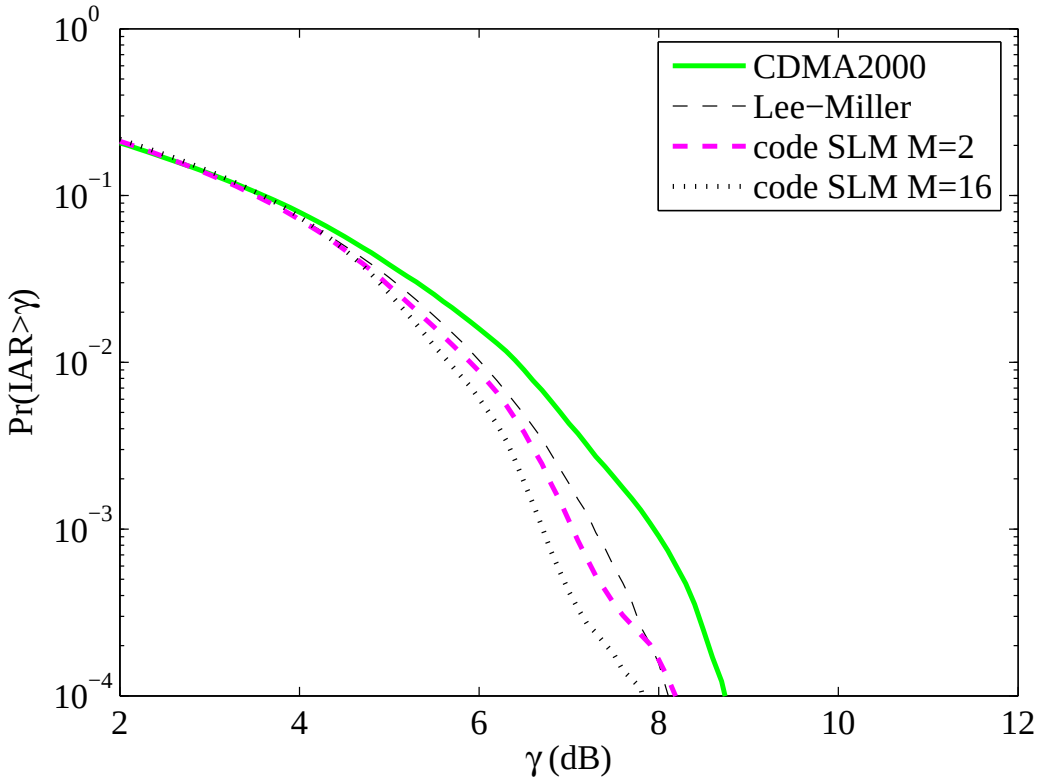


Figure 4.16: PAR reduction performance of Walsh code SLM for CDMA2000.

a performance - complexity tradeoff on the choice of N, and it should be decided according to the specific PAR reduction demand of the system.

4.3.4.4 PAR versus Number of Channels K

Fig. 4.18 shows the PAR versus the number of active channels (i.e., K) in a BSPNM-CDMA2000 system with $N = 1$. It is evident that the PAR increases as K increases. We can also see from Fig. 4.18 that the curves corresponding to varying M values are approximately parallel with each other when $K \geq 6$. This phenomena implies that the BSPNM has a consistent PAR reduction performance for various numbers of active channels.

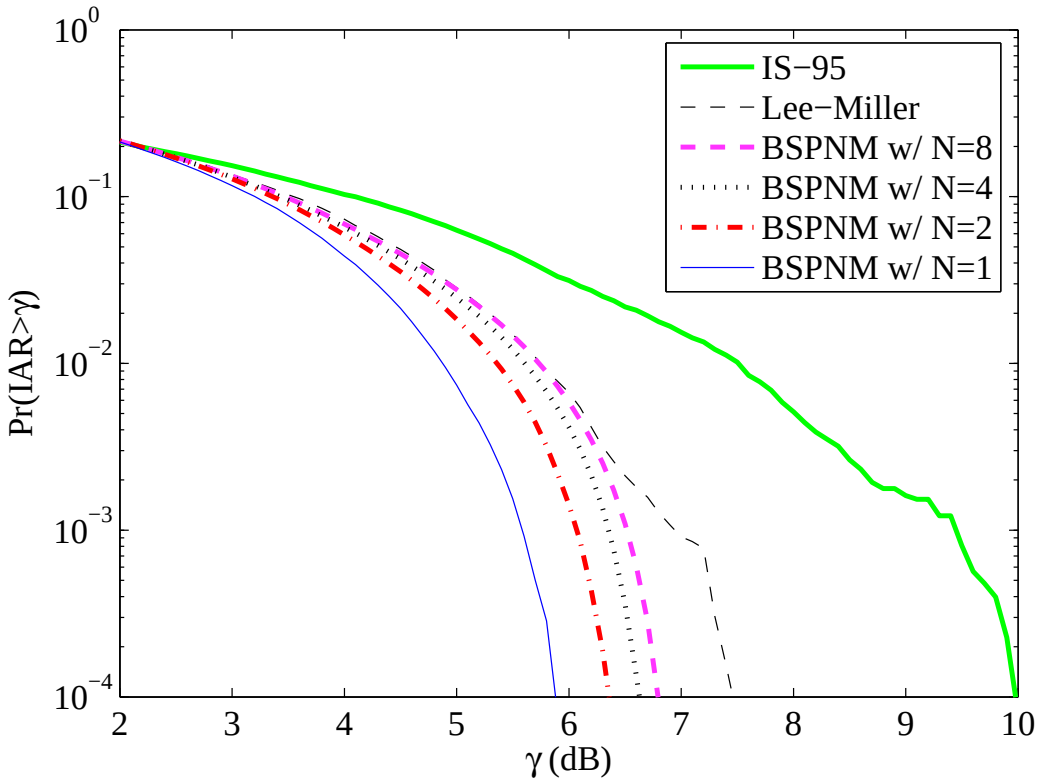


Figure 4.17: CCDF of the IAR of the BSPNM-CDMA2000 signal for different frame sizes.

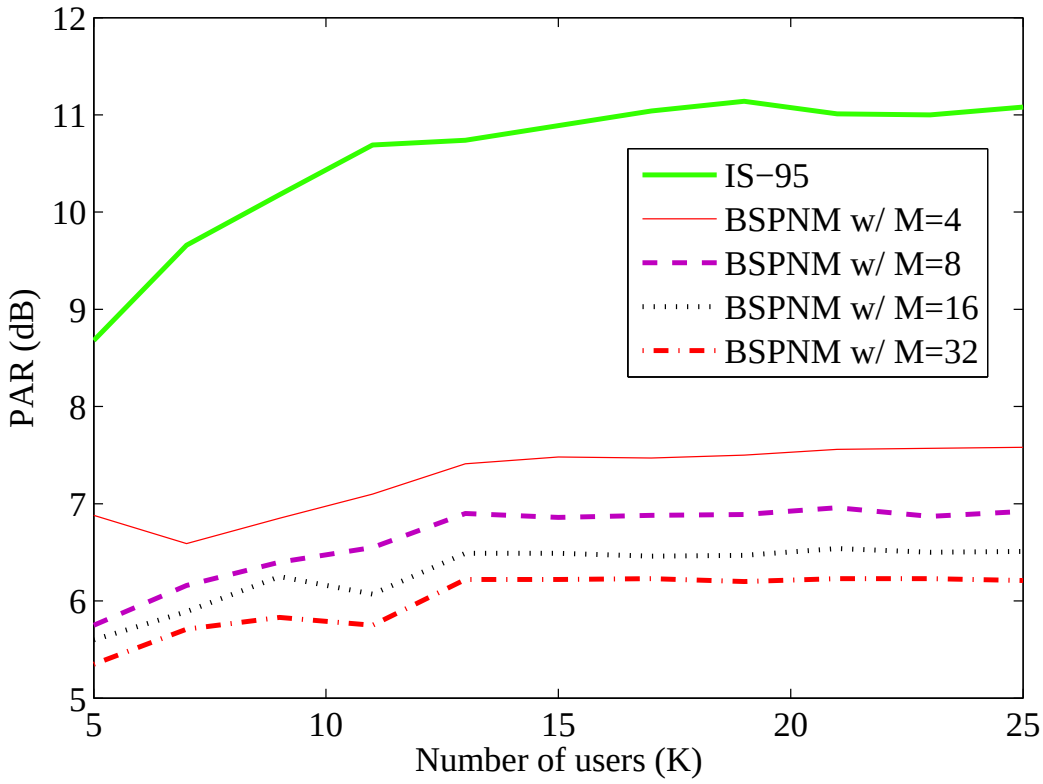


Figure 4.18: PAR of the BSPNM-CDMA2000 signal versus the number of active channels K. The PAR is measured at $\Pr\{\text{IAR} > \text{PAR}\} = 10^{-4}$.

CHAPTER 5

BLIND SLM WITH SUPERIMPOSED TRAINING

In Chapter 4, we proposed SLM algorithms to improve the power efficiency of OFDM (and forward link CDMA). To avoid the transmission of the side information in the SLM scheme for OFDM, we proposed to combine SLM with PTAM to facilitate joint channel estimation and PAR reduction; the resulting algorithm is called blind selected pilot tone modulation (BSPTM). However, PTAM requires dedicated pilot subcarriers for training, which consume valuable bandwidth. Thus, we are motivated to use the superimposed training scheme to achieve higher bandwidth efficiency. In this chapter, we will first analyze the PAR of the OFDM signal with superimposed training, and then modify the BSPTM algorithm to suit an OFDM system with superimposed training. We then obtain an effective solution which drastically improves both bandwidth efficiency and power efficiency of OFDM systems.

5.1 PAR Analysis for OFDM with Superimposed Training

In this section, we shall carry out probabilistic analysis of the PAR of the superimposed OFDM signal. Our main contributions are: i) derivation of the CCDF of the PAR of the superimposed OFDM signal, and ii) determination of the lower and upper bounds of the CCDF and the superimposed training sequences that achieve those bounds. We will show that by judiciously selecting the superimposed training sequence, we may sacrifice a little in the power that is dedicated to the information signal, but gain a lot in the power that is devoted to channel sounding, for the same amount of DC power consumed by a Class A or light Class AB PA. Such an "uneven" (and favorable) trade-off is possible because with a carefully designed superimposed training sequence, the average power of the transmitted superimposed OFDM signal can be increased without increasing the DC or peak power, rendering the PA as more efficiently utilized.

5.1.1 Superimposed Training for OFDM

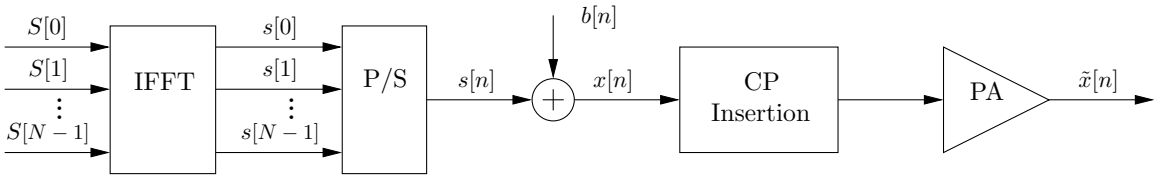


Figure 5.1: Discrete-time baseband OFDM transmitter for the i^{th} block with N subcarriers. The superimposed training sequence is $b[n]$; the baseband equivalent transmitted signal is $\tilde{x}[n]$.

Fig. 5.1 depicts the discrete-time baseband schematic of an OFDM transmission system with superimposed training. We next apply the superimposed training scheme described in Chapter 3 to the OFDM system.

Denote by $S[k]$ the kth sub-symbol (in the frequency domain) of an OFDM symbol with N subcarriers. The time-domain OFDM signal $s[n]$ is obtained by taking the normalized IDFT of $S[k]$:

$$s[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S[k] e^{j2\pi kn/N}.$$

Under the superimposed training framework [12, 51, 67, 77, 79], we add a known pilot sequence, denoted by $b[n]$, onto $s[n]$ to obtain $x[n] = s[n] + b[n]$. A length G cyclic prefix is then padded onto $x[n]$ and the resulting signal passes through a linear (or linearized) PA to yield

$$\tilde{x}[n] = \mathcal{G}\ x[n + N - G]_N, \quad 0 \le n \le N + G - 1, \tag{5.1}$$

where $\mathcal{G}$ is the linear gain of the PA, and $[n]_N$ is the residue of n divided by N. Without loss of generality, the PA gain can be assumed to be unity; i.e., $\mathcal{G} = 1$.

We consider a frequency-selective block fading channel model $h[n]$ with length $L \le G+1$. The received signal is

$$\tilde{y}[n] = \tilde{x}[n] * h[n] + v[n], \quad 0 \le n \le N + G - 1, \tag{5.2}$$

where $*$ denotes linear convolution; and $v[n]$ is the zero-mean stationary additive noise. After removing the cyclic prefix, the linear convolution becomes circular convolution (denoted

by $\circledast$), and we obtain

$$y[n] = x[n] \circledast h[n] + v[n] = b[n] \circledast h[n] + u[n] = d[n] + u[n], \quad 0 \le n \le N-1, \tag{5.3}$$

where $d[n] = b[n] \circledast h[n]$; and $u[n] = s[n] \circledast h[n] + v[n]$. It is evident that the cyclic prefix turns the linear convolution in (3.1) into the circular convolution in (5.3).

Similarly as in Chapter 3, we rewrite (5.3) into the matrix-vector format to obtain

$$\mathbf{y} = \mathbf{C}_b \mathbf{h} + \mathbf{u}, \tag{5.4}$$

where $\mathbf{C}_b$ is an $N \times L$ circulant matrix with $[b[0], b[1], \ldots, b[N-1]]^T$ in the first column, $\mathbf{y} = [y[0], y[1], \ldots, y[N-1]]^T$, $\mathbf{h} = [h[0], h[1], \ldots, h[L-1]]^T$, $\mathbf{u} = [u[0], u[1], \ldots, u[N-1]]^T$, and ^{T} denotes transpose. From (5.4), the LS CSI estimator is

$$\hat{\mathbf{h}} = {\mathbf{C}_b}^{\dagger} \mathbf{y}, \tag{5.5}$$

where ${\mathbf{C}_b}^{\dagger} = ({\mathbf{C}_b}^H \mathbf{C}_b)^{-1} {\mathbf{C}_b}^H$, and $\hat{\mathbf{h}}$ is the estimate of $\mathbf{h}$.

For a length-N $b[n]$, eq. (5.5) involves large matrix multiplications. We are thus motivated to consider periodic $b[n]$ sequences for which (5.5) can be implemented more efficiently.

5.1.1.1 Periodic Pilot Sequences

In [77], it was shown that for a fast time-varying frequency-selective (i.e., doubly-selective) channel, periodic pilot sequences enable the estimation of Doppler shifts. As we will see next, periodic pilot sequences can also simplify the implementation of (5.5) for the case of frequency-selective block fading channels.

When $b[n]$ is periodic with period P, we need $P \ge L$ in order to identify the $h[n]$ of length L. Assume for simplicity, that $R = N/P$ is an integer. Let us define a $P \times N$ averaging matrix $\mathbf{A} = \frac{1}{R}[\underbrace{\mathbf{I}_P, \ldots, \mathbf{I}_P}_{R}]$, where $\mathbf{I}_P$ is a $P \times P$ identity matrix. From (5.4), we obtain

$$\mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{C}_b \mathbf{h} + \mathbf{A}\mathbf{u}. \tag{5.6}$$

Now,

$$\bar{\mathbf{y}} \triangleq \mathbf{A}\mathbf{y}$$

is a $P \times 1$ vector whose ith element is

$$\bar{y}[i] = \frac{1}{R} \sum_{r=0}^{R-1} y[rP + i],\ 0 \leq i \leq P - 1,$$

or the ith "synchronized" average of $y[n]$. Similarly, $\bar{\mathbf{u}} \triangleq \mathbf{A}\mathbf{u}$ is a $P \times 1$ vector whose ith element is

$$\bar{u}[i] = \frac{1}{R} \sum_{r=0}^{R-1} u[rP + i].$$

Since $b[n]$ is periodic with period P, a synchronized average of $b[n]$ yields a single period of $b[n]$, denoted by $b_0[n]$; i.e., $b[n]$ can be seen as a periodic extension of $b_0[n]$.

Let us define a $P \times P$ circulant matrix

$$\mathbf{C}_{b_0} = \begin{bmatrix} b[0] & b[P-1] & \dots & b[1] \\ b[1] & b[0] & \dots & b[2] \\ & & \vdots & \\ b[P-1] & b[P-2] & \dots & b[0] \end{bmatrix}.$$

We can rewrite eq. (5.6) as

$$\bar{\mathbf{y}} = \mathbf{C}_{b_0}\mathbf{h} + \bar{\mathbf{u}}. \tag{5.7}$$

The LS estimator for $\mathbf{h}$ is

$$\hat{\mathbf{h}} = \mathbf{C}_{b_0}^{\dagger}\bar{\mathbf{y}} = (\mathbf{C}_{b_0}^{H}\mathbf{C}_{b_0})^{-1}\mathbf{C}_{b_0}^{H}\bar{\mathbf{y}}. \tag{5.8}$$

Since the size of the matrix $\mathbf{C}_{b_0}$ $(P \times P)$ is much smaller than that of the matrix $\mathbf{C}_b$ $(N \times N)$, eq. (5.8) is more computationally efficient than eq. (5.5). As mentioned before, the elements of $\bar{\mathbf{y}}$ can be found by simple averaging operations. Since $u[n]$ contains information about $h[n]$ as well, it is possible to improve the $\hat{h}[n]$ estimate of (5.8) using a weighted LS estimator by exploiting the covariance structure of $u[n]$ [77] or operating (5.8) and the subsequent symbol detector in a decision feedback mode [43].

5.1.1.2 Symbol Detection

We can rewrite (5.3) as $y[n] = (s[n] + b[n]) \circledast h[n] + v[n],\ 0 \leq n \leq N - 1$. Therefore, once $h[n]$ is estimated, a one-tap frequency-domain equalizer can be applied on each subcarrier to yield

$$\hat{S}[k] = \frac{Y[k]}{\hat{H}[k]} - P[k], \quad 0 \leq k \leq N - 1, \tag{5.9}$$

where $\hat{H}[k] = \sum_{n=0}^{L-1} \hat{h}[n]e^{-j\frac{2\pi kn}{N}}$, $Y[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} y[n]e^{-j\frac{2\pi kn}{N}}$, and $P[k]$ is defined similarly to $Y[k]$. $\hat{S}[k]$ is then passed through a minimum-distance classifier to yield the symbol estimates $\bar{S}[k]$.

In this section, we will show that with superimposed training in OFDM, it is possible to reduce the PAR of the signal that goes through the PA by judiciously selecting the $b[n]$ sequence. If we consider the DC power $\mathcal{P}_{dc}$ drawn by the PA as fixed, we can select a $b[n]$ such that the average transmit power ($\mathcal{P}_t = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2$) of the superimposed OFDM signal is larger than the $\mathcal{P}_t$ of the OFDM signal alone. Therefore, some diversion of $\mathcal{P}_t$ to $b[n]$ can be offset by a larger increase in $\mathcal{P}_t$ itself since a more efficient region of the PA is used. Our power analysis will provide insight on the choice of the optimal $b[n]$ from the PAR perspective. We shall make this point by first examining the distribution of the PAR under the superimposed training framework.

5.1.2 Distribution of the PAR

Strictly speaking, the continuous-time PAR of (2.3) is a more accurate indicator of the power efficiency of the PA. However, it is very challenging to derive the distribution of the PAR for a superimposed OFDM signal. Since our major goal is to investigate the effect of different pilot sequences on the resulting PARs, the relative changes of the PARs are of interest. It has been shown in [62] that there is a consistent gap between the CCDF of the continuous-time PAR and that of the Nyquist-rate sampled PAR, we will focus on deriving the CCDF of PAR($x[n]$).

The PAR of an OFDM symbol $\{x[n]\}_{n=0}^{N-1}$ should be defined as

$$\text{PAR}(x[n]) = \frac{\max\limits_{0\le n\le N-1}\{|x[n]|^2\}}{\frac{1}{N}\sum_{n=0}^{N-1}|x[n]|^2}. \tag{5.10}$$

Since for N large [39],

$$\frac{1}{N}\sum_{n=0}^{N-1}|x[n]|^2 \to \bar{E}[|x[n]|^2] \triangleq \frac{1}{N}\sum_{n=0}^{N-1}E[|x[n]|^2],$$

we adopt the following PAR definition when $x[n]$ is generally (non)stationary:

$$\text{PAR}(x[n]) = \frac{\max\limits_{0\le n\le N-1}\{|x[n]|^2\}}{\bar{E}[|x[n]|^2]}. \tag{5.11}$$

The quantity in (5.11) is easier to handle since the denominator is now a constant. Ochiai [47] evaluated the distribution of the so-called symbol-wise PAR in (5.10) using a numerical method. In order to obtain a closed-form expression for the CCDF of the PAR with superimposed training, we assume N large and stick to the PAR definition in (5.11). Note that we do not consider the cyclic prefix in the PAR analysis since it does not affect the peak or the average power [62].

5.1.2.1 CCDF of PAR for Superimposed OFDM

Assume without loss of generality that the frequency-domain OFDM signal $\{S[k]\}_{k=0}^{N-1}$ is i.i.d., drawn from a known constellation with variance σ_s^2. When the number of subcarriers N is large, it follows from the Central Limit Theorem that the time-domain OFDM signal $\{s[n]\}_{n=0}^{N-1}$ is approximately i.i.d. complex Gaussian distributed with zero-mean and variance σ_s^2 [62,63]. Hence, $x[n] = s[n]+b[n]$ is independent complex Gaussian distributed with time-varying mean $b[n]$ and variance σ_s^2.

Denote by $r[n]$ the instantaneous power of $x[n]$; i.e., $r[n] = |x[n]|^2$. We infer that $r[n]$ has a noncentral χ^2 distribution (see e.g., [52]) with two degrees of freedom and noncentrality parameter $|b[n]|^2$. The cumulative distribution function (CDF) of $r[n]$ is

$$F_{R_n}(r) = 1 - e^{-\frac{|b[n]|^2+r}{\sigma_s^2}} \sum_{k=0}^{\infty} (\frac{|b[n]|}{\sqrt{r}})^k I_k(\frac{|b[n]|\sqrt{r}}{\sigma_s^2/2}), \tag{5.12}$$

where $I_k(a)$ is the kth-order modified Bessel function of the first kind, which can be represented by the infinite series

$$I_k(a) = \sum_{n=0}^{\infty} \frac{(a/2)^{2n+k}}{n!\ (n+k)!}, \tag{5.13}$$

for $a \geq 0$.

The complementary CDF (CCDF) of the PAR of $x[n]$ is given by

$$\begin{aligned} \Pr\{\text{PAR} > \gamma\} &= 1 - \Pr\left\{ \frac{\max\limits_{0\leq n\leq N-1}\{r[n]\}}{\bar{E}[r[n]]} \leq \gamma \right\} \\ &= 1 - \Pr\{r[n] \leq \gamma \bar{E}[r[n]],\ 0 \leq n \leq N-1\}. \end{aligned}$$

Since $r[n]$ at different n's are independent, the probability on the right hand side (RHS) of the above equation can be expressed in a product form.

Proposition 5.1 (CCDF of PAR). *Suppose that $s[n]$ is a time-domain OFDM signal with N subcarriers and average power σ_s^2, $b[n]$ is a known pilot sequence with average power $\sigma_p^2 = \frac{1}{N}\sum_{n=0}^{N-1}|b[n]|^2$, and $x[n] = s[n] + b[n]$. The CCDF of the PAR of $x[n]$ is given by*

$$\Pr\{\text{PAR} > \gamma\} = 1 - \prod_{n=0}^{N-1} F_{R_n}\{\gamma\ (\sigma_s^2 + \sigma_p^2)\}, \tag{5.14}$$

where $F_{R_n}(r)$ is the CDF of the noncentral χ^2 distribution given in (5.12).

Proposition 1 implies that the distribution of the PAR of $x[n]$ depends on the number of subcarriers N, the average power of the information signal σ_s^2, the average power of the pilot sequence σ_p^2, as well as the magnitude of the superimposed pilot sequence $|b[n]|$.

We shall limit our discussions to periodic pilot sequences. An aperiodic $b[n]$ can be viewed as a special periodic $b[n]$ with period $P = N$. As we have alluded to in Section 5.1.1.1, the periodic $b[n]$ can be beneficial for simplifying the LS estimate for the frequency-selective block fading channel, or to track the Doppler frequencies of the doubly-selective channel. The idea of exploiting the cyclostationarity in $y[n]$ induced by the periodic $b[n]$ has been discussed in [12, 51, 67, 77, 79].

Because the noncentrality parameter is $|b[n]|^2$ and $b[n]$ is periodic, the CDF $F_{R_n}(r)$ in (5.12) is also periodic in n. The CCDF of $x[n]$ in (5.14) can thus be simplified to:

$$\Pr\{\text{PAR} > \gamma\} = 1 - \left(\prod_{n=0}^{P-1} F_{R_n}\{\gamma\ (\sigma_s^2 + \sigma_p^2)\}\right)^R, \tag{5.15}$$

where for simplicity, we have assumed that $R = N/P$ is an integer.

Substituting (5.12) and (5.13) into (5.15), we write out the expression for the CCDF of the PAR as

$$\Pr\{\text{PAR} > \gamma\} = 1 - e^{-\frac{N\gamma}{\sigma_s^2/(\sigma_s^2+\sigma_p^2)}} \left(\prod_{n=0}^{P-1} \sum_{k=1}^{\infty} \frac{(\frac{\gamma}{\sigma_s^2/(\sigma_s^2+\sigma_p^2)})^k}{k!} \frac{\Gamma(k, |b[n]|^2/\sigma_s^2)}{(k-1)!}\right)^R, \tag{5.16}$$

where

$$\Gamma(n, a) = (n-1)!e^{-a}\sum_{l=0}^{n-1}\frac{a^l}{l!} \tag{5.17}$$

is the incomplete gamma function [25].

Since the PAR in (5.14) depends on the specific $|b[n]|$, we discuss next the types of $b[n]$ that will lead to the worst and the best PARs, respectively. In the sequel, we shall consider the power allocation factor β (c.f. (3.16)) in superimposed training as fixed.

5.1.2.2 Worst-case PAR

Under the superimposed training framework, the periodic impulse sequence,

$$b_1[n] = \sqrt{P}\sigma_p \sum_{l=-\infty}^{+\infty} \delta(n - lP - n_0),\ 0 \le n_0 \le P-1,\ 0 \le n \le N-1, \tag{5.18}$$

has been shown in [12,79] to yield extremely simple channel estimates. Our next proposition shows that such a pilot sequence is not ideal however, from the PAR perspective.

Proposition 5.2 (Worst-case PAR). *Suppose that the power allocation factor $\beta = \sigma_p^2/(\sigma_s^2 + \sigma_p^2)$ is fixed. Among all periodic sequences $b[n]$ of period P, the impulse sequence in (5.18) gives rise to the worst PAR for the superimposed OFDM signal $x[n] = s[n] + b[n]$; i.e., eq. (5.15) is upper bounded by the following, which is the CCDF of the PAR for $x_1[n] = s[n] + b_1[n]$:*

$$\Pr\{\text{PAR} > \gamma\} = 1 - e^{-\frac{R\gamma}{1-\beta}}(1 - e^{-\frac{\gamma}{1-\beta}})^{N-R} \left(\sum_{k=1}^{\infty} \frac{(\frac{\gamma}{1-\beta})^k}{k!} \frac{\Gamma(k, \frac{\beta}{1-\beta})}{(k-1)!} \right)^R. \tag{5.19}$$

Proof: See Appendix A.

As an example, let us consider $\sigma_s^2 = 2$, $N = 128$, $P = 8$, and $R = N/P = 16$. When $b_1[n]$ of (5.18) is used, the theoretical CCDFs (solid lines) of the PAR of $x_1[n] = s[n] + b_1[n]$ is shown in Fig. 5.2 for various power allocation factors β. The empirical CCDFs (dash-dotted lines) were obtained by averaging over 10^7 OFDM symbols. It is interesting to observe that the CCDF curves corresponding to $\beta = 0$ (corresponding to the original OFDM signal) and $\beta = 0.9$ cross over, indicating that at a low CCDF level (e.g., 10^{-4}), the PAR can actually be reduced even with the periodic impulse sequence $b_1[n]$. However, for a fixed β, *Proposition 5.2* indicates that $b_1[n]$ will result in the worst-case PAR among all periodic $b[n]$ sequences with the same period.

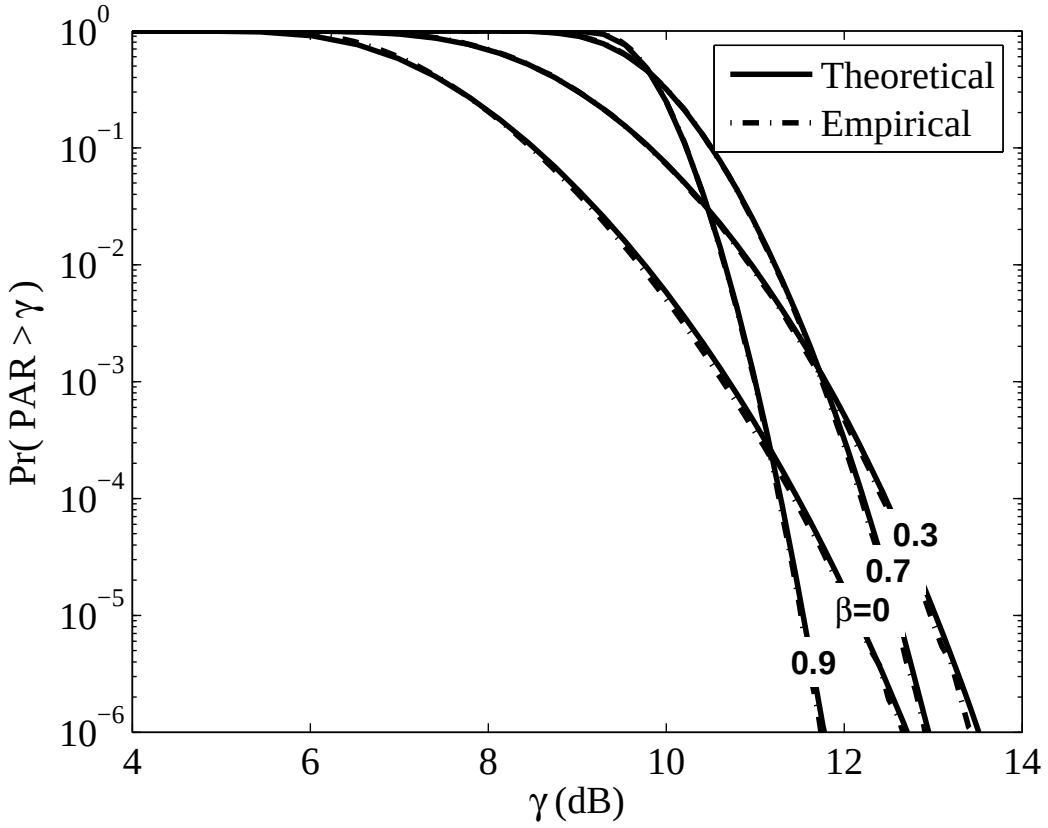


Figure 5.2: CCDF of PAR of $x_1[n] = s[n] + b_1[n]$ with $N = 128$ and $P = 8$.

5.1.2.3 Best-case PAR

It was shown in [67, 77] and [51] that a periodic $b[n]$ with a constant magnitude results in low MSE for the channel estimate and excellent receiver BER performance. We further advocate periodic $b[n]$ sequences with constant $|b[n]|$ from a PAR perspective.

Proposition 5.3 (Best-case PAR). *Suppose that the power allocation factor $\beta = \sigma_p^2/(\sigma_s^2 + \sigma_p^2)$ is fixed. Among all periodic sequences $b[n]$ of period P, the sequence $b_2[n]$ with constant magnitude $|b_2[n]| = \sigma_p$, $\forall\ n$, gives rise to the best PAR for the superimposed OFDM signal $x[n] = s[n] + b[n]$; i.e., eq. (5.15) is lower-bounded by the following, which is the CCDF of the PAR for $x_2[n] = s[n] + b_2[n]$:*

$$\Pr\{\text{PAR} > \gamma\} = 1 - e^{-\frac{N\gamma}{1-\beta}} \left(\sum_{k=1}^{\infty} \frac{(\frac{\gamma}{1-\beta})^k}{k!} \frac{\Gamma(k, \frac{\beta}{1-\beta})}{(k-1)!} \right)^N . \tag{5.20}$$

Proof: See Appendix A.

Proposition 5.3 may be intuitive, since the constant magnitude $b_2[n]$ has the lowest possible PAR of 1 (i.e., 0 dB). However, proving that (5.20) is the lower bound of (5.15) is not trivial, since there is not a simple expression linking the PAR of $x[n] = s[n] + b[n]$ to that of $s[n]$ and $b[n]$.

When $\beta = 0$; i.e., $b[n] = 0$, the CCDF expressions in (5.19) and (5.20) both reduce to

$$\Pr\{\text{PAR} > \gamma\} = 1 - (1 - e^{-\gamma})^N, \tag{5.21}$$

which is commonly cited in the literature [47, 62] as the CCDF of the PAR of the OFDM signal $s[n]$.

As an example, consider $\sigma_s^2 = 2$, $N = 128$, $P = 8$, and $R = N/P = 16$. When $|b_2[n]|$ is constant, the CCDF of the PAR of $x_2[n] = s[n] + b_2[n]$ is shown in Fig. 5.3 for various power allocation factors β. We can see that the CCDF curve lowers as β is increased.

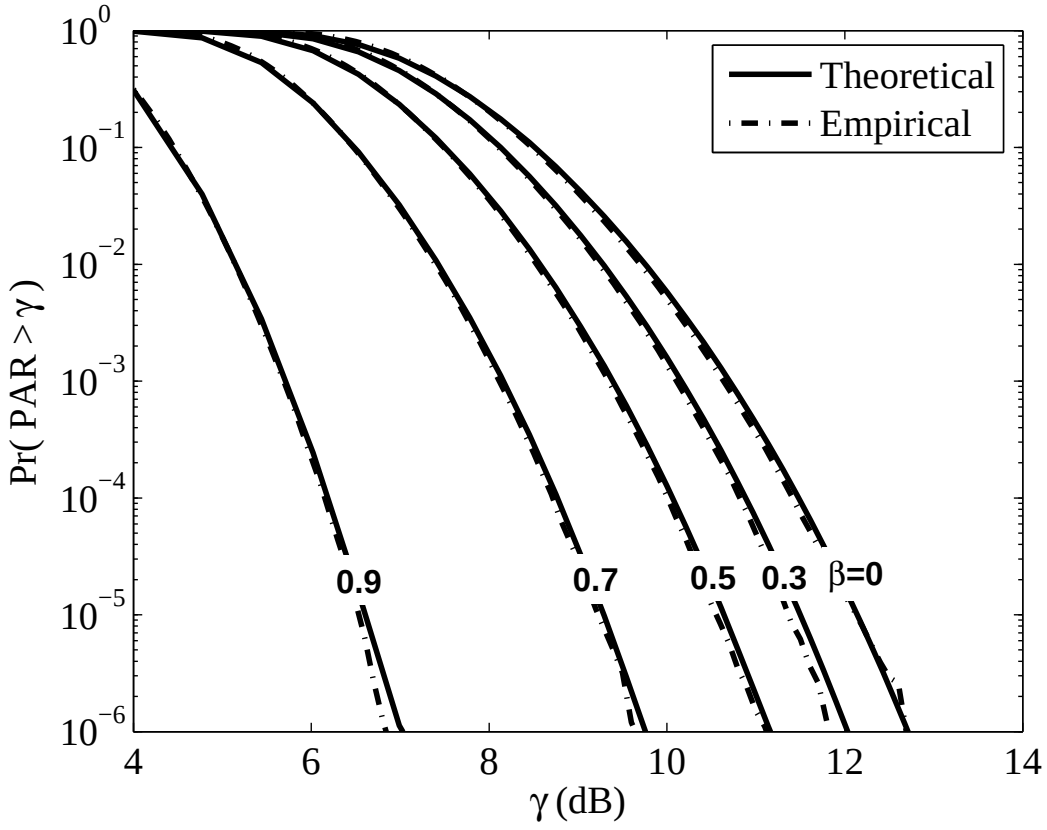


Figure 5.3: CCDF of PAR of $x_2[n] = s[n] + b_2[n]$ with $N = 128$ and $P = 8$.

Fig. 5.2 and Fig. 5.3 demonstrate the accuracy of our PAR analysis as the differences between the theoretical analysis and the simulation results are negligible.

It was shown in [3,50] that the optimal way to place pilot tones in PTAM for frequency-selective block fading channels is to modulate P pilots with equal power onto equally spaced subcarriers (N/P subcarriers apart). Taking the IDFT of such PTAM-OFDM data, we find that the corresponding time-domain signal also fits into the superimposed training framework: $x[n] = s[n] + b[n]$, where $s[n]$ is the time-domain OFDM signal with periodic null subcarriers in the frequency domain, and $b[n]$ is periodic with period P. The existence of P null subcarriers in $S[k]$ in PTAM is the main difference between PTAM and superimposed training. When $P \ll N$, the existence of the null subcarriers can be neglected and thus the PAR analysis on superimposed training can be applied to PTAM as well.

Since it is possible to reduce the PAR with superimposed training in OFDM, we next describe methods to realize power savings under the superimposed training framework.

5.1.3 Power Analysis

Fig. 5.4 shows the input-output characteristic of an ideal linear PA. To exploit the maximum efficiency of the PA, the maximum output power should reach $\mathcal{P}_{\text{sat}}$ (the saturation point of the PA). The so-called output backoff (OBO) is the ratio between $\mathcal{P}_{\text{sat}}$ and the average output power (denoted by $\mathcal{P}_t$) of the PA; the so-called input backoff (IBO) is the ratio between $\mathcal{P}_{\text{lin}}$ and the average input power of the PA. Although OBO and IBO are generally different, for a linear PA and in the absence of clipping, IBO=OBO. When the maximum input power

$$\max_{0 \le n \le N-1} |x[n]|^2 = \mathcal{P}_{\max}, \tag{5.22}$$

we have IBO=PAR. The efficiency of the PA is defined as [16]

$$\eta = \frac{\mathcal{P}_t}{\mathcal{P}_{dc}}. \tag{5.23}$$

Class A PAs are the most linear with efficiency given by $\eta = 0.5/\text{OBO}$ [16]. Since we have established that OBO=IBO=PAR for an ideal linear Class A PA with input satisfying (5.22), we infer that

$$\eta = \frac{0.5}{\text{PAR}}. \tag{5.24}$$

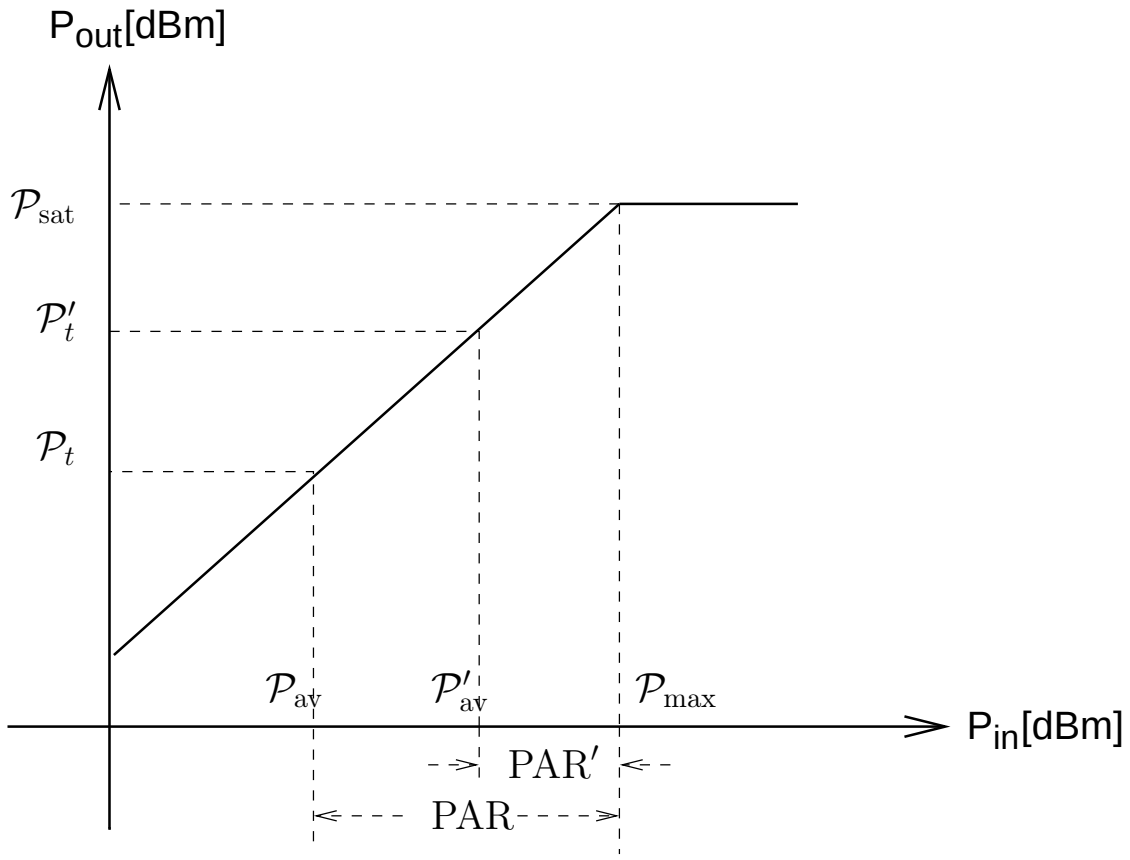


Figure 5.4: Output power vs. input power for a linearized PA, or a PA that obeys the ideal linear model.

Combining (5.23) and (5.24), we obtain

$$\mathcal{P}_{dc} = 2\mathcal{P}_t \cdot \text{PAR}, \tag{5.25}$$

or equivalently,

$$\mathcal{P}_t = \frac{0.5\mathcal{P}_{dc}}{\text{PAR}}. \tag{5.26}$$

We shall consider the following two scenarios, and assess the impact of the PAR distribution on the power tradeoffs:

i. If the average transmit power $\mathcal{P}_t$ is fixed;

ii. If the DC power (e.g., from the battery) $\mathcal{P}_{dc}$ is fixed.

When $\mathcal{P}_t$ is fixed, we assume that the transmission PA can be re-biased according to any PAR change to ensure the most efficient utilization of the PA. Taking the expected value

Table 5.1: $E[\text{PAR}(s[n] + b_2[n])]$ corresponding to different number of subcarriers and various power allocation factors.

N	$\beta = 0$	$\beta = 0.1$	$\beta = 0.3$	$\beta = 0.6$	$\beta = 0.9$
64	4.80	4.76	4.49	3.70	2.25
128	5.45	5.41	5.08	4.10	2.42
256	6.14	6.09	5.67	4.50	2.58
512	6.84	6.76	6.25	4.92	2.73
1024	7.50	7.43	6.84	5.32	2.88

Table 5.2: $E[\frac{1}{\text{PAR}(s[n]+b_2[n])}]$ corresponding to different number of subcarriers and various power allocation factors.

N	$\beta = 0$	$\beta = 0.1$	$\beta = 0.3$	$\beta = 0.6$	$\beta = 0.9$
64	0.22	0.22	0.23	0.28	0.45
128	0.19	0.19	0.21	0.25	0.42
256	0.17	0.17	0.18	0.23	0.39
512	0.15	0.15	0.16	0.21	0.37
1024	0.14	0.14	0.15	0.19	0.35

on both sides of (5.25), we infer that

$$E[\mathcal{P}_{dc}] = 2\mathcal{P}_t \cdot E[\text{PAR}]. \tag{5.27}$$

Therefore, the average DC power consumption is proportional to the expected value of the PAR, which can be evaluated numerically using the CCDF expression in (5.15).

Table 5.1 shows various $E[\text{PAR}]$ values when $|b[n]|$ is constant. The PAR reducing capability of superimposed training is evident. Take for example, $N = 128$,

$$\frac{E[\text{PAR}]|_{\beta=0.6}}{E[\text{PAR}]|_{\beta=0}} = \frac{4.10}{5.45} = 75\%, \tag{5.28}$$

implying that a 25% of savings in DC power is possible when $b_2[n]$ with $\beta = 0.6$ is used with OFDM.

On the other hand, when $\mathcal{P}_{dc}$ is fixed, $E[\mathcal{P}_t]$ is proportional to $E[\frac{1}{\text{PAR}}]$ according to (5.26). Fixing $\mathcal{P}_{dc}$ is equivalent to fixing $\mathcal{P}_{\text{sat}}$ for Class A ($\mathcal{P}_{\text{sat}} = 2\mathcal{P}_{dc}$) or light Class AB PAs. Linear scaling has been proposed in [73] and [47] to fully exploit the efficiency of the PA by keeping the peak power fixed.

Table 5.2 shows various $E[\frac{1}{\text{PAR}}]$ values when $|b[n]|$ is constant. Continue with the above

example for $N = 128$, we infer that

$$\frac{E[\frac{1}{\text{PAR}}]|_{\beta=0.6}}{E[\frac{1}{\text{PAR}}]|_{\beta=0}} = \frac{0.25}{0.19} = 131\%, \tag{5.29}$$

implying that a 31% of average transmit power increase is possible when $b_2[n]$ of $\beta = 0.6$ is used with OFDM.

In the PAR reduction literature, attention has been paid to peak power reduction methods. We argue that since the PA is peak power limited, and its efficiency is determined by the average output power, it makes sense to investigate PAR reduction methods that aim at increasing the average power while keeping the peak power fixed.

5.1.4 Simulations

In this section (unless otherwise specified), the number of sub-carriers in each OFDM block is $N = 128$, the period of $b[n]$ is $P = 8$, thus the number of periods is $R = N/P = 16$. The OFDM symbols $\{S[k]\}_{k=0}^{N-1}$ were drawn from an i.i.d. 4QAM constellation. We consider two pilot sequences:

$$b_1[n] = \sqrt{P}\sigma_p \sum_{l=0}^{R-1} \delta(n - lP), \quad 0 \le n \le N-1, \tag{5.30}$$

$$b_2[n] = \begin{cases} \sigma_p \; e^{j\frac{\pi}{P}([n]_P^2 + [n]_P)}, & P \text{ odd}, \; 0 \le n \le N-1, \\ \sigma_p \; e^{j\frac{\pi}{P}([n]_P^2 + 2[n]_P)}, & P \text{ even}, \; 0 \le n \le N-1. \end{cases} \tag{5.31}$$

$b_1[n]$ is a periodic impulse sequence, and $b_2[n]$ is a periodic chirp sequence (thus it has a constant magnitude) [51, 77]. The Fourier transforms of both sequences have equi-spaced and equi-powered profiles in the frequency-domain; i.e., both are optimal training sequences from the channel estimation perspective [50, 51].

5.1.4.1 Bounds on the CCDF of the PAR

According to *Proposition 5.2* and *Proposition 5.3*, the CCDF of the PAR of $x[n] = s[n]+b[n]$ is upper bounded by that of $x[n]$ with $b_1[n]$ and lower bounded by that of $x[n]$ with $b_2[n]$, for the same period P and power allocation factor β. For $\beta = 0.5$, Fig. 5.5 shows the CCDFs of the PAR of $s[n]$ (dotted line), that of $x[n]$ with $b_1[n]$ (dash-dotted line), that of $x[n]$ with $b_2[n]$ (dashed line), and that of $x[n]$ with a random $b[n]$ (solid line) for the same P and β.

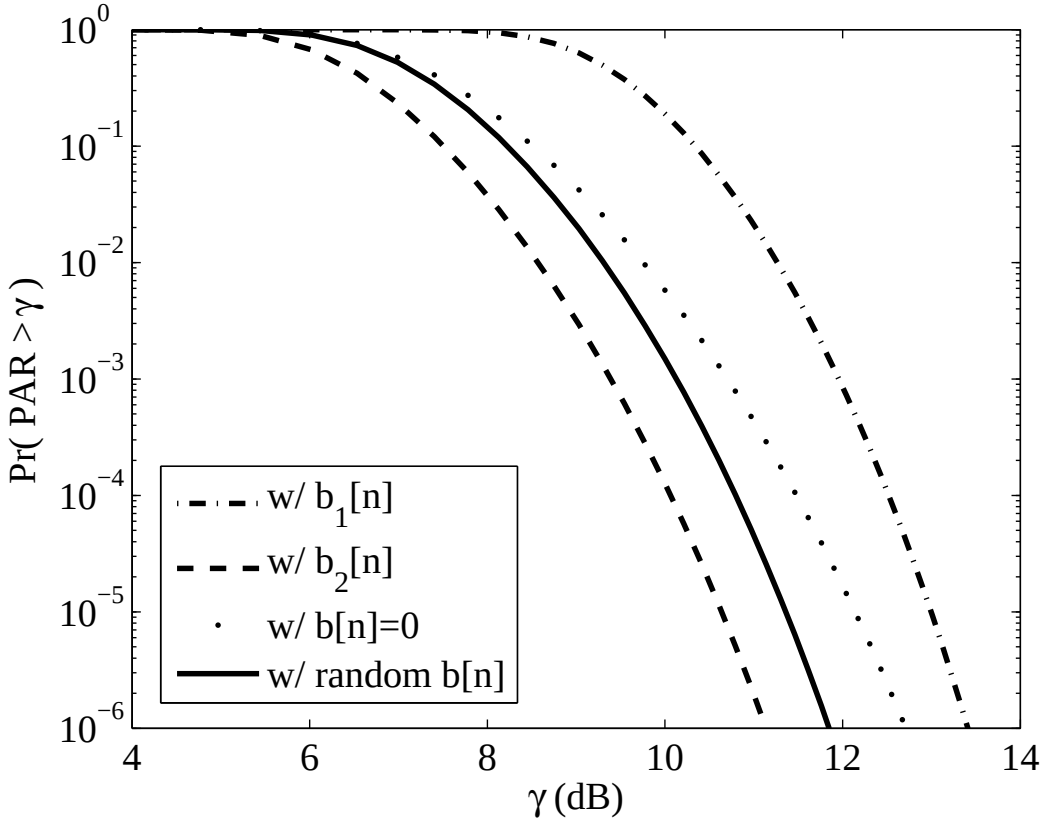


Figure 5.5: CCDF of PAR of $x[n] = s[n] + b[n]$: the upper bound is reached with $b[n] = b_1[n]$, and the lower bound is reached with $b[n] = b_2[n]$.

We observe from Fig. 5.5 that the CCDF curve corresponding to the randomly selected $b[n]$ falls in between those curves corresponding to $b_1[n]$ and $b_2[n]$ (this was always the case in our simulations), thus illustrating *Propositions 5.2* and 5.3.

5.1.4.2 Average Output Power and Power Allocation

Here we assume that the peak power of the PA, $\mathcal{P}_{\text{sat}}$, is fixed (or equivalently, $\mathcal{P}_{dc}$ is fixed), and linear scaling [47,73] is applied to adjust the average transmit power $\mathcal{P}_t$ according to the PAR. In Fig. 5.6, we show $E[\mathcal{P}_t]$ versus β for superimposed OFDM signals with $b[n] = b_1[n]$ and $b[n] = b_2[n]$, for the OFDM signal alone (i.e., $b[n] = 0$), and for PTAM-OFDM signals with $b[n] = b_1[n]$ and $b[n] = b_2[n]$, respectively. 10^4 OFDM symbols were used to obtain the empirical averages. From Fig. 5.6, we observe that when compared with $E[\mathcal{P}_t]$ of the OFDM only case (dash-dotted line), $b_2[n]$ caused an increase in $E[\mathcal{P}_t]$, $b_1[n]$ caused a decrease in

$E[\mathcal{P}_t]$, for both superimposed training and PTAM. Interestingly, for the same β, $E[\mathcal{P}_t]$ is almost the same for PTAM-OFDM and superimposed OFDM. With $b_2[n]$, several dBs of average transmit power increase can be achieved.

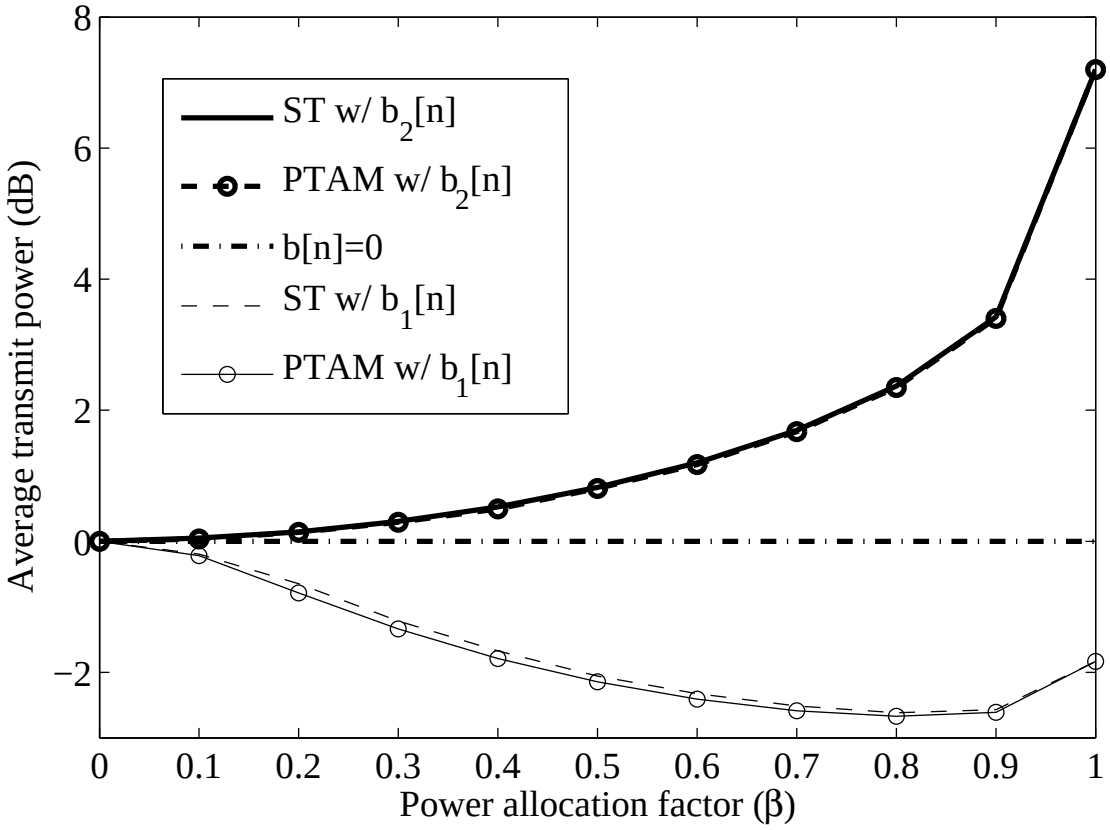


Figure 5.6: Average transmit power $E[\mathcal{P}_t]$ versus β for superimposed training (ST) with $b[n] = b_1[n]$ (dashed line) and $b[n] = b_2[n]$ (thick solid line), for PTAM with $b[n] = b_1[n]$ (solid line with circles) and $b[n] = b_2[n]$ (thick dashed line with circles), and for OFDM with $b[n] = 0$ (dash-dotted line), when $N = 128$ and $P = 8$.

5.1.4.3 BER Performance

In this example, we compare the BER performance of OFDM with superimposed training to that obtained when PTAM is used, when a one-tap frequency-domain equalizer is employed. For superimposed training, the channel estimator (5.8), the symbol estimator (5.9), and the subsequent symbol classifier operated in a decision feedback mode with one iteration [43]. The frequency-selective block fading channel is modeled by an FIR filter $h[n]$ with length $L = 8$ and unit energy (i.e., $\sum_{n=0}^{L-1} E|h[n]|^2 = 1$); $h[n]$ is fixed over a given block, but varies from block to block. The tap coefficients were generated from a zero-mean complex

Table 5.3: Indoor test environment tap-delay line profile.

Tap	1	2	3	4	6	8
Relative delay (nsec)	0	100	200	300	500	700
Average power (dB)	0	-3.6	-7.2	-10.8	-18	-25.2

Gaussian distribution with power delay profile shown in Table 5.3 [2], which simulates a typical in-door propagation environment. The OFDM data rate was 10M symbols/sec, i.e., 20 Mbps for a 4QAM modulation. $1,000$ independent Monte Carlo runs were performed. The SNR is defined as in (4.33).

Fig. 5.7 shows the BER performances for the case with $N = 128,\ P = 8$. With regard to the power allocation factor β, we used $\beta = 0.3$ (empirically determined to be "optimal") for superimposed training and $\beta = 0.2$ (which is optimal according to [50, eq. (24)]) for PTAM. The reason that the two optimal β's are different is that it takes more pilot power to overcome the equivalent additive noise $u[n] = s[n] \circledast h[n] + v[n]$ (c.f. (5.3)) in superimposed training than to overcome $v[n]$ in PTAM. A cyclic prefix of length $G = L-1 = 7$ corresponds to a bandwidth efficiency of $\frac{N}{N+G} = 94.8\%$ for superimposed training, but $\frac{N-L}{N+G} = 88.9\%$ for PTAM. From Fig. 5.7, we can see that the BER performance with $b[n] = b_2[n]$ outperformed the case with $b[n] = b_1[n]$ for both superimposed training and PTAM. This is because the judicious choice of $b_2[n]$ traded a small amount of information signal power $\mathcal{P}_s = \mathcal{P}_t(1-\beta)$ for much more training pilot power $\mathcal{P}_t\beta$. The BER performance of superimposed training is inferior to that of PTAM, as a price paid for its higher bandwidth efficiency.

Fig. 5.8 shows the BER performances with $N = 512$ (all other parameters are the same as in the previous example). Comparing Fig. 5.7 with Fig. 5.8, we can see that with the larger N, the performance of both superimposed training and PTAM approached that of the known channel case. However, when we consider $\mathcal{P}_{dc}$ as fixed, the BER may not always decrease with N. The reasons are: (i) when N is larger, the PAR tends to be larger for OFDM, thus SNRe $\propto 1/\text{PAR}$ becomes smaller, which has a negative impact on the BER; (ii) when N is larger and β is fixed, channel estimation performance improves under both superimposed training and PTAM, having a positive impact on the BER performance [22,50,51]. Because of the opposite effects of (i) vs. (ii), when N is increased, the overall BER may or may not

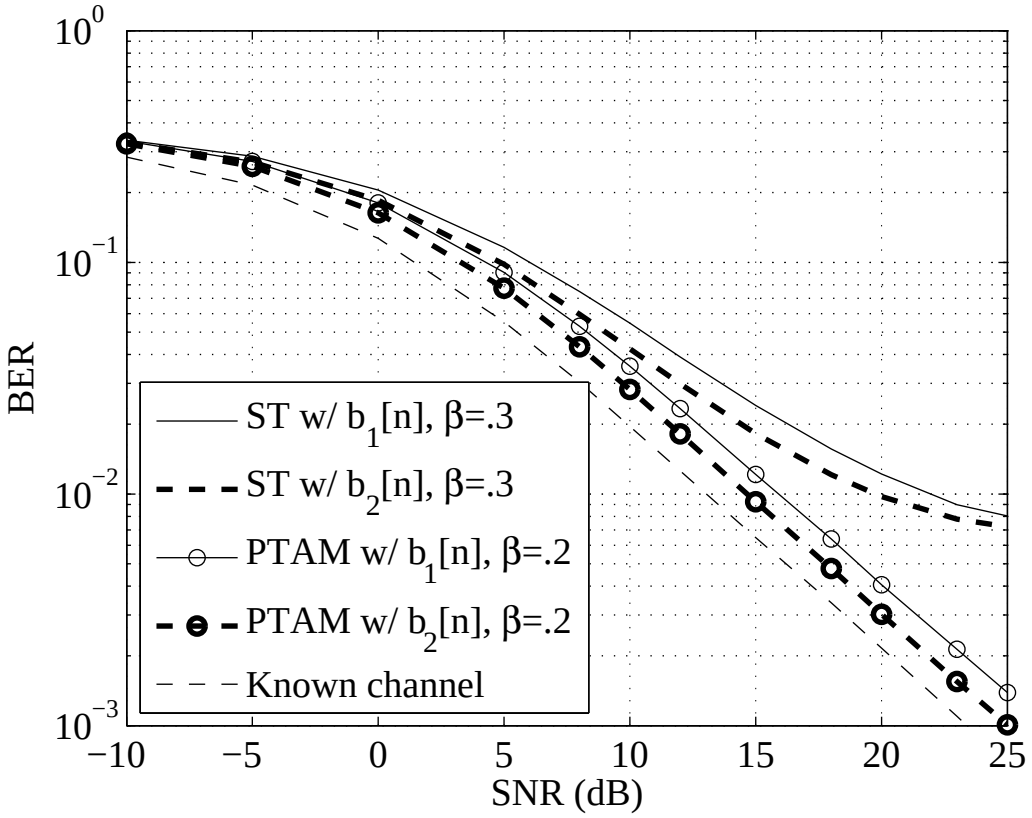


Figure 5.7: BER performances with $N = 128$, $P = 8$. $b_2[n]$ outperforms $b_1[n]$ in both superimposed training (ST) and PTAM cases.

decrease, even for the known channel case.

If the channel does not change within a number of blocks and this information is taken into account in the channel estimation stage, then the BER performance will improve for both superimposed training and PTAM; the improvement in the former is more significant than the improvement in the latter.

The above two examples demonstrate that superimposed training achieves a higher bandwidth efficiency than PTAM, but at a cost of BER performance degradation. For both training techniques, $b_2[n]$ is preferred over $b_1[n]$ since the former has better BER performance due to lower PAR values of the resulting transmitted OFDM signals.

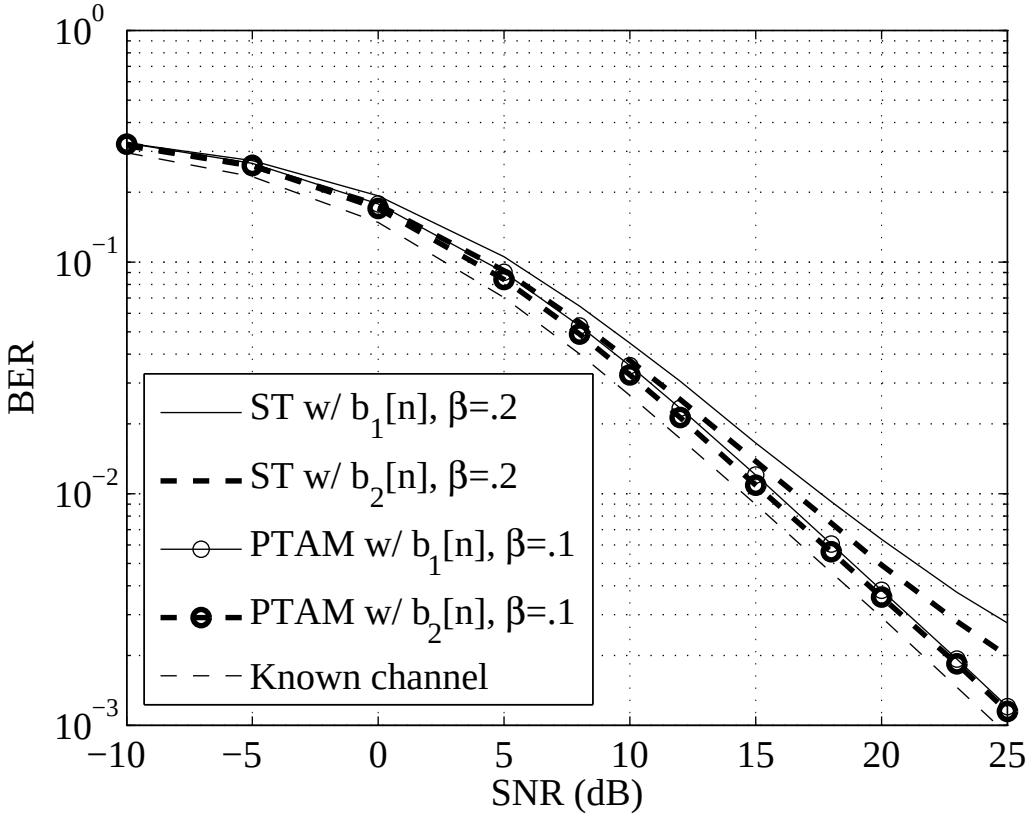


Figure 5.8: BER performances with $N = 512$, $P = 8$. $b_2[n]$ outperforms $b_1[n]$ in both superimposed training (ST) and PTAM cases.

5.2 BSPTM with Superimposed Training

Capitalizing on the BSPTM algorithm in Chapter 4, we propose to replace PTAM with superimposed training, the aim being to improve the bandwidth efficiency as well as to improve the power efficiency. We adopt superimposed training for the purpose of channel estimation; i.e., we add pilot tones onto the data. We can achieve a higher bandwidth efficiency this way, but the CSI estimate will not be as accurate as in the dedicated training case. We then propose two methods to improve the CSI estimate. For better power efficiency, we employ the SLM technique to reduce the PAR, and embed the side information about the SLM index in the position of the superimposed pilot tones. We demonstrate the PAR reducing capability of the proposed method as well as the resulting BER performance. In comparison with PTAM, our proposed method is advantageous at medium to low SNRs (< 10 dB) in the sense that higher bandwidth efficiency can be achieved without sacrificing the BER.

5.2.1 PTAM versus Superimposed Training

We have reviewed PTAM-OFDM in Section 4.2.1; we have also described superimposed training-OFDM in Section 5.1.1. In this section, we derive a generalized superimposed training framework, in which both superimposed training and PTAM can be shown as special cases. We adopt the following notations. A time-domain sequence is represented by a lower case letter; e.g., $s[n]$. Its frequency-domain representation is denoted by the corresponding capital case letter; e.g., $S[k]$. A bold-faced lower case letter denotes a vector of the frequency-domain sequence; e.g., $\mathbf{s} = [S[0],\ S[1],\ \ldots,\ S[N-1]]^T$. Its time-domain counterpart is denoted by adding a subscript $_t$; e.g., $\mathbf{s}_t = [s[0],\ s[1],\ \ldots,\ s[N-1]]^T$. A bold-faced capital case letter denotes a matrix.

In OFDM, a length N frequency-domain subsymbol $\mathbf{s} = [S[0],\ S[1],\ \ldots,\ S[N-1]]^T$ is transformed into the time-domain to yield the transmitted signal as

$$\mathbf{s}_t = \mathbf{F}^H\ \mathbf{s},$$

where $\mathbf{F}$ is the $N \times N$ normalized discrete Fourier transform (DFT) matrix with $\mathbf{F}(k,n) = 1/\sqrt{N}\ e^{-j2\pi kn/N}$, and $\mathbf{F}^H$ is its conjugate transpose. In PTAM-OFDM, P pilot tones are

inserted in the frequency domain in order to acquire the CSI; $P \geq L$ is assumed where L is the length of the finite impulse response (FIR) channel. The frequency-domain transmitted signal is

$$\mathbf{x} = \mathbf{s} + \mathbf{b}, \tag{5.32}$$

where $\mathbf{s}$ consists of $N-P$ information subsymbols (data) and P zeros, $\mathbf{b} = [B[0],\ B[1],\ \ldots,\ B[N-1]]^T$ consists of P pilot tones and $N-P$ zeros. The locations of the zeros in $\mathbf{s}$ and $\mathbf{b}$ are carefully selected so that when $S[k] = 0,\ B[k] \neq 0$, and when $B[k] = 0,\ S[k] \neq 0$. The placement of the pilot tones and the allocation of the power between $\mathbf{b}$ and $\mathbf{s}$ have been studied in [3, 46, 50]. The same problem has also been investigated for single carrier systems [22]. All these studies point to the same optimality conditions for the pilot tones for estimating frequency selective block fading channels.

According to [50], the optimal placement strategy is to modulate $P = L$ pilot tones with equal power onto equi-spaced subcarriers. For simplicity, let us assume that the number of sub-carriers N is an integer multiple of P; i.e., $R = N/P$ is an integer. Next define a set of P equi-spaced pilot tone indices as

$$\mathbf{\Omega}_0 \triangleq \Big\{ k_i \ \Big|\ k_i = iR + \theta_0, 0 \leq i \leq P-1, 0 \leq \theta_0 \leq R-1 \Big\}, \tag{5.33}$$

which can be characterized by the shift parameter θ_0 alone.

We consider a frequency selective block fading channel, which is modeled by a time-invariant (over each OFDM block) FIR filter $\mathbf{h} = [h[0],\ h[1],\ \ldots,\ h[L-1]]^T$. With the insertion/removal of the cyclic prefix (CP), the frequency selectivity of the channel appears flat on each subcarrier; i.e.,

$$\mathbf{y} = \mathbf{D}_H\ \mathbf{x} + \mathbf{v}, \tag{5.34}$$

where $\mathbf{y} = [Y[0],\ Y[1],\ \ldots,\ Y[N-1]]^T$ is the received signal after the CP removal and the application of the DFT, $\mathbf{D}_H$ is a diagonal matrix with the channel DFT coefficients; i.e.,

$$H[k] = \sum_{n=0}^{L-1} h[n] e^{-j2\pi kn/N},\ 0 \leq k \leq N-1$$

on the diagonal, and $\mathbf{v} = [V[0],\ V[1],\ \ldots,\ V[N-1]]^T$ is the normalized DFT of the additive noise $v[n]$; i.e.,

$$V[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} v[n] e^{-j2\pi kn/N}.$$

Substituting (5.32) into (5.34), we obtain

$$\begin{aligned} \mathbf{y} &= \mathbf{D}_H\ (\mathbf{s} + \mathbf{b}) + \mathbf{v} \\ &= \mathbf{D}_H\ \mathbf{s} + \mathbf{D}_B\ \mathbf{F}_{1:L}\ \mathbf{h} + \mathbf{v}, \end{aligned} \tag{5.35}$$

where $\mathbf{D}_B$ is a diagonal matrix with $\mathbf{b}$ on the diagonal, and $\mathbf{F}_{1:L} = \sqrt{N}\mathbf{F}(:, 1:L)$ is an $N \times L$ DFT matrix consisting of the first L columns of $\sqrt{N}\mathbf{F}$.

Let $\mathbf{B} = \mathbf{D}_B\ \mathbf{F}_{1:L}$ and

$$\mathbf{B}^\dagger = (\mathbf{F}_{1:L}^H \mathbf{D}_B^H \mathbf{D}_B \mathbf{F}_{1:L})^{-1}\ \mathbf{F}_{1:L}^H \mathbf{D}_B^H \tag{5.36}$$

be the pseudoinverse of $\mathbf{B}$. Since $\mathbf{B}$ has full column rank, we can pre-multiply $\mathbf{B}^\dagger$ on both sides of (5.35) to obtain

$$\mathbf{B}^\dagger\ \mathbf{y} = \mathbf{B}^\dagger\ \mathbf{D}_H\ \mathbf{s} + \mathbf{h} + \mathbf{B}^\dagger\ \mathbf{v}. \tag{5.37}$$

Since the pilot tones and the data are decoupled in PTAM, we have $\mathbf{D}_B^H\ \mathbf{s} = \mathbf{0}$, which allows us to write $\mathbf{D}_B^H \mathbf{D}_H\ \mathbf{s} = \mathbf{0}$ as well. Thus, the first term on the RHS of (5.37) is zero and

$$\mathbf{B}^\dagger\ \mathbf{y} = \mathbf{h} + \mathbf{B}^\dagger\ \mathbf{v}. \tag{5.38}$$

We infer from (5.38) that the LS estimate of the channel $\mathbf{h}$ is

$$\hat{\mathbf{h}} = \mathbf{B}^\dagger\ \mathbf{y}. \tag{5.39}$$

In other words, $\hat{\mathbf{h}} = \mathbf{h} + \mathbf{B}^\dagger\mathbf{v}$; and hence the unknown $\mathbf{s}$ does not affect the accuracy in $\hat{\mathbf{h}}$.

Afterwards, the data can be estimated as

$$\hat{\mathbf{s}} = \mathbf{D}_{\hat{H}}^{-1}\ \mathbf{y} - \mathbf{b}, \tag{5.40}$$

where $\mathbf{D}_{\hat{H}}^{-1}$ is a diagonal matrix with $1/\hat{H}[k]$ on the diagonal[1]. $\hat{\mathbf{s}}$ is then decoded to yield the $\bar{\mathbf{s}}$ estimate belonging to the symbol constellation.

[1]To use this one-tap equalizer, it is assumed that $\hat{H}[k] \neq 0,\ \forall\ k$. To cope with this, typically the symbol stream is coded across subcarriers. The minimum MSE estimator can also be used.

The inserted pilot tones in PTAM reduce the bandwidth efficiency. To avoid the bandwidth efficiency loss, we can resort to superimposed training. As we discussed in Chapter 3 and Section 5.1.1.1, a periodic pilot sequence $b[n]$ with period P is commonly used in superimposed training. The Fourier transform of this periodic time-domain pilot sequence turns out to be P equi-spaced pilot tones. The only difference between PTAM and superimposed training is whether the pilot tones are "inserted" or "superimposed" in the frequency domain.

5.2.2 Generalized Superimposed Training

Note that with superimposed training, the pilot tones are no longer decoupled from the data; i.e., $\mathbf{D}_B^H \mathbf{s} \neq \mathbf{0}$. Thus, (5.38) does not hold any more. Since $E[\mathbf{D}_H\ \mathbf{s}] = \mathbf{0}$, equation (5.39) can still be used to estimate the CSI, although the performance of the channel estimate will not be as accurate as in the PTAM case. To improve (5.39) for the superimposed training case, we next propose a generalized superimposed training framework that allows us to control the amount of distortion from $\mathbf{s}$.

Under the superimposed training framework, it can be shown (e.g., [51, 80]) that the variance of the channel estimate in (5.39) increases with the information signal power σ_s^2, which is undesirable. Therefore, we can reduce the variance of $\hat{h}[l]$ by lowering the average power of those data subsymbols that are co-located with the pilot tones. For example, let us set

$$X[k] = \begin{cases} \sqrt{\alpha} S[k] + B[k], & k \in \mathbf{\Omega}_0 \\ S[k], & k \in \mathbf{\Omega}_0^{\perp}, \end{cases} \tag{5.41}$$

i.e.,

$$E[|S[k]|^2] = \begin{cases} \alpha \sigma_s^2, & k \in \mathbf{\Omega}_0 \\ \sigma_s^2, & k \in \mathbf{\Omega}_0^{\perp}, \end{cases} \tag{5.42}$$

where $\mathbf{\Omega}_0^{\perp}$ denotes the complement of $\mathbf{\Omega}_0$ and $0 \leq \alpha \leq 1$. $\alpha = 0$ corresponds to PTAM, whereas $\alpha = 1$ corresponds to conventional superimposed training. Let us denote by β, the power allocation factor, which is the ratio between the total power allocated to the pilot

tones and the total transmitted power. For example, in PTAM,

$$\beta = \frac{P\sigma_p^2}{P\sigma_p^2 + (N-P)\sigma_s^2}, \tag{5.43}$$

whereas for superimposed training with (5.42),

$$\beta = \frac{P\sigma_p^2}{P\sigma_p^2 + [N-(1-\alpha)P]\sigma_s^2}, \tag{5.44}$$

with $\sigma_p^2 = 1/P\sum_{k\in\mathbf{\Omega}_0}|B[k]|^2$ and $\sigma_s^2 = E[|S[k \in \mathbf{\Omega}_0^{\perp}]|^2]$. Comparing with the power allocation factor defined in (3.16), we can see that eq. (5.44) redefines β for superimposed training in the frequency domain.

5.2.2.1 Improved Channel Estimate

We can explore the CSI contained in $\mathbf{D}_H\mathbf{s}$ as described below. We infer from (5.35)

$$\mathbf{u} = \mathbf{y} - \mathbf{D}_B\mathbf{F}_{1:L}\mathbf{h} = \mathbf{D}_H\mathbf{s} + \mathbf{v}. \tag{5.45}$$

If $S[k]$ has a constant modulus; i.e., $|S[k]| = \sigma_s$, we can take the magnitude squared on both sides of (5.45) to obtain

$$|\mathbf{u}|^2 = |\mathbf{h}_f|^2\sigma_s^2 + |\mathbf{v}|^2 + 2\mathcal{R}\{\mathbf{D}_H\mathbf{D}_s\mathbf{v}^*\},$$

where $|\cdot|^2$ is an element-wise magnitude squared operator for a scalar, a vector, or a matrix, $\mathbf{h}_f$ is the diagonal of $\mathbf{D}_H$, $\mathbf{D}_s$ is a diagonal matrix with $\mathbf{s}$ on the diagonal, $\mathcal{R}\{\cdot\}$ denotes the real part, and $(\cdot)^*$ denotes complex conjugation. Treating the last two items in $|\mathbf{u}|^2$ as noise and (5.39) as the initial estimate of $h[l]$, we can estimate the magnitude response of $h[l]$ as

$$|\hat{\mathbf{h}}_f| = \sqrt{\frac{|\mathbf{y} - \mathbf{D}_B\mathbf{F}_{1:L}\hat{\mathbf{h}}|^2 - \sigma_v^2}{\sigma_s^2}}, \tag{5.46}$$

where again, the $\sqrt{\cdot}$ operation is carried out on each element. We propose to take the magnitude response from (5.46) but take the phase response from (5.39) to form an improved CSI estimate; i.e.,

$$\hat{H}[k] = |\hat{H}_2[k]|\ e^{j\angle\hat{H}_1[k]}, \quad 0 \le k \le N-1, \tag{5.47}$$

where $\hat{H}_1[k]$ is the DFT of the $\hat{h}[l]$ in (5.39), and $|\hat{H}_2[k]|$ comes from the kth element of $|\hat{\mathbf{h}}_f|$ in (5.46).

Knowing that the channel has length $L < N$, the CSI estimate in (5.47) can be further improved by smoothing $\hat{H}[k]$. Specifically, we take the N-point IFFT of $\hat{H}[k]$, truncate the resulting $\hat{h}[l]$ to retain only the first L samples, and subsequently take the FFT of the truncated $\hat{h}[l]$ to obtain its DFT coefficients for use in the equalizer.

5.2.2.2 Pilot Tone Selection

The main difference between superimposed training and PTAM is that superimposed training has both data and pilots on the pilot subcarriers, but PTAM has only pilots on those subcarriers. Both schemes use the CSI estimator in (5.39) and they have the same optimality condition for the placement of the pilot tones, i.e., equi-spaced and equi-powered pilot tones [50, 51]. Among such optimal $B[k]$'s, two classes are of interest; i.e.,

$$B_1[k] = \begin{cases} \sigma_p \, e^{j\phi_p}, & k \in \mathbf{\Omega}_0, \\ 0, & k \in \mathbf{\Omega}_0^{\perp}, \end{cases} \qquad 0 \le k \le N-1, \tag{5.48}$$

$$B_2[k] = \mathrm{DFT}\{b_2[n]\}, \tag{5.49}$$

where the time-domain sequence $b_2[n]$ has a constant-magnitude and is periodic with period P. The general rules for constructing $B_2[k]$ sequences (equi-powered in both time and frequency domains and periodic in time/equi-spaced in frequency) are not clear, but the following periodic chirp sequence is a solution [11, 51]:

$$b_2[n] = \begin{cases} \sigma_p \, e^{j\frac{\pi}{P}([n]_P^2 + [n]_P)}, & P \text{ odd}, \ 0 \le n \le N-1, \\ \sigma_p \, e^{j\frac{\pi}{P}([n]_P^2 + 2[n]_P)}, & P \text{ even}, \ 0 \le n \le N-1, \end{cases} \tag{5.50}$$

where $[n]_P$ is the residue of n divided by P. We will elaborate next on the use of $B_1[k]$ and $B_2[k]$.

When $B[k] = B_1[k]$, the CSI estimate in (5.39) can be greatly simplified. By substituting (5.48) into (5.39), we infer that $\mathbf{D}_B^H \mathbf{D}_B = \sigma_p^2 \, \mathbf{I}_{\mathbf{\Omega}_0}$, where $\mathbf{I}_{\mathbf{\Omega}_0}(k,k) = 1$ for $k \in \mathbf{\Omega}_0$ and zero otherwise. In this case, the first term on the RHS of (5.36) becomes a scaled identity matrix $(\mathbf{F}_{\mathbf{\Omega}_0,1:L}^H \, \mathbf{F}_{\mathbf{\Omega}_0,1:L})^{-1}/\sigma_p^2 = \mathbf{I}/(P\sigma_p^2)$, where $\mathbf{F}_{\mathbf{\Omega}_0,1:L}$ is a replica of $\mathbf{F}_{1:L}$ with the kth

row, $k \in \mathbf{\Omega}_0^{\perp}$, being zeroed out. Therefore, (5.39) can be rewritten as

$$\begin{aligned}\hat{\mathbf{h}} &= \frac{1}{P\sigma_p \exp(j\phi_p)}\, \mathbf{F}^H_{\mathbf{\Omega}_0,1:L}\, \mathbf{F}\, \mathbf{y}_t \\ &= \frac{1}{P\sigma_p \exp(j\phi_p)/\sqrt{N}} \frac{1}{R} \left[e^{-j2\pi\theta_0 \frac{0}{R}}\mathbf{I},\ e^{-j2\pi\theta_0 \frac{1}{R}}\mathbf{I},\ \ldots,\ e^{-j2\pi\theta_0 \frac{R-1}{R}}\mathbf{I}\right] \mathbf{y}_t, \quad (5.51)\end{aligned}$$

or in scalar form,

$$\hat{h}[l] = \frac{1}{P\sigma_p e^{j\phi_p}} \frac{1}{R} \sum_{r=0}^{R-1} y[rP+l]\ e^{-j2\pi\theta_0 \frac{l}{R}}, \quad 0 \le l \le L-1, \quad (5.52)$$

which is the phase-shifted synchronized average of the received time-domain signal $y[n]$. The receiver is very simple since the channel estimate (5.52) only uses a first-order statistic, and the equalizer is a simple one-tap equalizer which is an appealing feature of OFDM.

The normalized IDFT of $B_1[k]$ is

$$b_1[n] = \frac{P}{\sqrt{N}}\sigma_p\ e^{j\phi_p} \sum_{l=0}^{R-1} \delta[n-lP]\ e^{j2\pi\theta_0 \frac{l}{R}}, \quad (5.53)$$

which is a periodic impulse pilot sequence. As we show in Section 5.1.2, among all periodic sequences with the same period, superimposed training with $b_1[n]$ gives rise to the worst case PAR whereas superimposed training with $b_2[n]$ (c.f. (5.50)) leads to the best case PAR. The gap between the CCDF of the best case PAR and that of the worst case PAR is a function of the number of subcarriers N, the number of pilot tones P and the power allocation factor β. For example, when $N = 128$, $P = 8$ and $\beta = 0.5$, the PAR of the transmitted OFDM signal with $b_1[n]$ is 2 dB higher than that of the OFDM signal with $b_2[n]$ at the clipping probability of 10^{-4}.

We will show in Section 5.2.4.1 that when SLM is employed to reduce the PAR of the transmitted signal, $B_1[k]$ and $B_2[k]$ do not differ much in their impact on the final PAR value after SLM. Table 5.2.2.2 compares the $B_1[k]$ and $B_2[k]$ in terms of PAR and CSI estimation complexity.

Table 5.4: Comparison of two types of pilot tones: $B_1[k]$ vs. $B_2[k]$. TD and FD stand for time domain and frequency domain, respectively.

	Periodic	Periodic Impulse	Periodic Constant Magnitude
$b[n]$ (TD)	$b[n]_P$	$c_1 \sum_{l=0}^{R-1} \delta[n-lP]\ e^{j\phi_l}$	$\begin{cases} \sigma_p\ e^{j\frac{\pi}{P}([n]_P^2+[n]_P)}, & P \text{ odd} \\ \sigma_p\ e^{j\frac{\pi}{P}([n]_P^2+2[n]_P)}, & P \text{ even} \end{cases}$
$B[k]$ (FD)	$\big\lvert B[k \in \mathbf{\Omega}_0]\big\rvert = c$	$B_1[k] = \begin{cases} \sigma_p\ e^{j\phi_p}, & k \in \mathbf{\Omega}_0 \\ 0, & k \in \mathbf{\Omega}_0^{\perp} \end{cases}$	$\big\lvert B_2[k \in \mathbf{\Omega}_0]\big\rvert = c$
$\hat{\mathbf{h}}$	$(\mathbf{C}_b{}^H\mathbf{C}_b)^{-1}\mathbf{C}_b{}^H\mathbf{y}_t$	$\frac{1}{c_1 R}\ \left[\mathbf{I}\ e^{-j\phi_0}, \ldots, \mathbf{I}\ e^{-j\phi_{R-1}}\right]\mathbf{y}_t$	$(\mathbf{C}_{b_0}^H\mathbf{C}_{b_0})^{-1}\mathbf{C}_{b_0}^H\bar{\mathbf{y}}_t$
Complexity	$\mathcal{O}\{\max\{P^3, NP\}\}$	$\mathcal{O}\{P\}$	$\mathcal{O}\{P^3\}$
PAR	—	Highest	Lowest

5.2.3 Blind Selected Mapping with Superimposed Training

It is straightforward to modify the BSPTM technique proposed in Section 4.2.2 for the superimposed training framework. Among the M equivalent representations, the mth superimposed training OFDM signal is given by

$$X^{(m)}[k] \;=\; \begin{cases} \sqrt{\alpha}S[k] + B[k], & k \in \mathbf{\Omega}_m, \\ S[k], & k \in \mathbf{\Omega}_m^{\perp}, \end{cases} \tag{5.54}$$

for $0 \le k \le N-1$ and $0 \le m \le M-1$, where $\mathbf{\Omega}_m$ is represented by $\theta_0^{(m)}$ similar to the way that $\mathbf{\Omega}_0$ is represented by θ_0.

Next perform the phase rotations,

$$Z^{(m)}[k] = X^{(m)}[k]e^{j\phi^{(m)}[k]}. \tag{5.55}$$

Similar to SLM, $z^{(m)}(t)$ and $\mathrm{PAR}\left(z^{(m)}(t)\right)$ are evaluated and $z^{(\bar{m})}(t)$, which has the lowest PAR among $\{z^{(m)}(t)\}$, is transmitted. In other words, the optimum pilot tone location - phase sequence index is

$$\bar{m} = \underset{0 \le m \le M-1}{\operatorname{argmin}} \left\{\mathrm{PAR}\left(z^{(m)}(t)\right)\right\}. \tag{5.56}$$

The expected power of $X^{(m)}[k]$ is

$$E[|X^{(m)}[k]|^2] = \begin{cases} \alpha\sigma_s^2 + \sigma_p^2, & k \in \mathbf{\Omega}_m \\ \sigma_s^2, & k \in \mathbf{\Omega}_m^{\perp}. \end{cases} \tag{5.57}$$

Since $\sigma_p^2 \gg \sigma_s^2$ [3, 14, 50], we see that there is a disparity in the average power at the pilot tone locations vs. at the non-pilot tone locations. Since the phase rotations in (5.55) do not affect the power, the same average power profile holds for $Z^{(m)}[k]$.

The received frequency-domain signal is

$$Y[k] = Z^{(\bar{m})}[k]H[k] + V[k]. \tag{5.58}$$

As in Section 4.2.2.4, we first form the synchronized average of the instantaneous power of $Y[k]$ as

$$\rho[r] = \frac{1}{P} \sum_{l=0}^{P-1} |Y[lR + r]|^2. \tag{5.59}$$

At the receiver, we determine the optimum index $\bar{m}$, i.e., the actual pilot shift $\theta_0^{(\bar{m})}$, through peak detection in $\rho[r]$; i.e.,

$$\hat{\theta}_0^{(\bar{m})} = \underset{r \in \boldsymbol{\theta}}{\operatorname{argmax}} \{\rho[r]\}. \tag{5.60}$$

Since the receiver knows the set of possible values for $\mathbf{\Theta}$, from $\hat{\theta}_0^{(\bar{m})}$, a simple lookup table search yields $\hat{\bar{m}}$.

The accuracy in $\hat{\bar{m}}$ is critical for the decoding of $\hat{S}[k]$ at the receiver. If $\hat{\bar{m}}$ is erroneous for one particular OFDM block, the BER will be high for that block. The success of the algorithm largely hinges upon ensuring $\sigma_p^2 \gg \sigma_s^2$. This as well as the finite alphabet nature of $\mathbf{\Theta}$ help to make (5.60) robust over frequency selective channels. Similar findings for the PTAM-SLM case were presented in Section 4.2.2.4.

5.2.4 Simulations

In the examples in this section, we assume that the number of sub-carriers $N = 128$, the length of the FIR channel $L = 4$, and the number of pilot tones $P = L = 4$. According to [50], the optimum power allocation factor for PTAM for the above parameters is $\beta = 0.15$.

For superimposed training, we chose $\beta = 0.3$ because it yields relatively good performance for both low and high SNR scenarios. The weighting factor in (5.54) was chosen to be $\alpha = 0.5$.

The data were independently drawn from a QPSK constellation with Gray coding. The SNR is defined as in (4.33). Again, for simplicity, we can have $\phi^{(m)}[k]$ i.i.d. drawn from $\{0, \pi\}$ with equal probability so that $\exp(j\phi^{(m)}[k]) = \pm 1$ with equal probability. As such, (5.55) can be implemented without multiplications but with selected sign changes only.

5.2.4.1 PAR Reduction Performance

In this example, we approximate the continuous-time PAR by evaluating the discrete-time PAR of the 4-time oversampled OFDM signal [62]. 10^6 independent Monte Carlo trials were conducted.

Fig. 5.9 shows the empirical CCDF curves (solid lines) of the PAR of the transmitted signal $z^{(\bar{m})}(t)$ for different number of selections, M, along with the CCDF of the PAR of the original OFDM signal and that of the PTAM-OFDM signal. We can see from the figure that the PTAM-OFDM signal had a higher PAR than the OFDM signal when $B_1[k]$ was used, but a lower PAR when $B_2[k]$ was used. By employing SLM, we can significantly reduce the PAR. We observe that when $M = 8$, the proposed algorithm achieved approximately 3.5 dB of PAR reduction (compared with the original OFDM signal) at the CCDF level of 10^{-4} with either $B_1[k]$ or $B_2[k]$. We also see from Fig. 5.9 that the larger the M, the smaller the resulting PAR, and the smaller the gap between the CCDF corresponding to $B_1[k]$ and that corresponding to $B_2[k]$. On the other hand, the computational complexity of SLM increases as M increases. There is also a diminishing return in the PAR reduction capability as M further increases. As a rule of thumb, we recommend to use $\min\{R, 4\} \leq M \leq \min\{R,\ 8\}$.

5.2.4.2 BER Performance

Fig. 5.10 compares the BER performance of our proposed scheme with that of PTAM-OFDM for $N = 128$, $P = 4$, $\beta = 0.15$, and $M = 8$. When SLM is employed, $B_1[k]$ is used because of the corresponding low complexity in the CSI estimation. Otherwise, $B_2[k]$

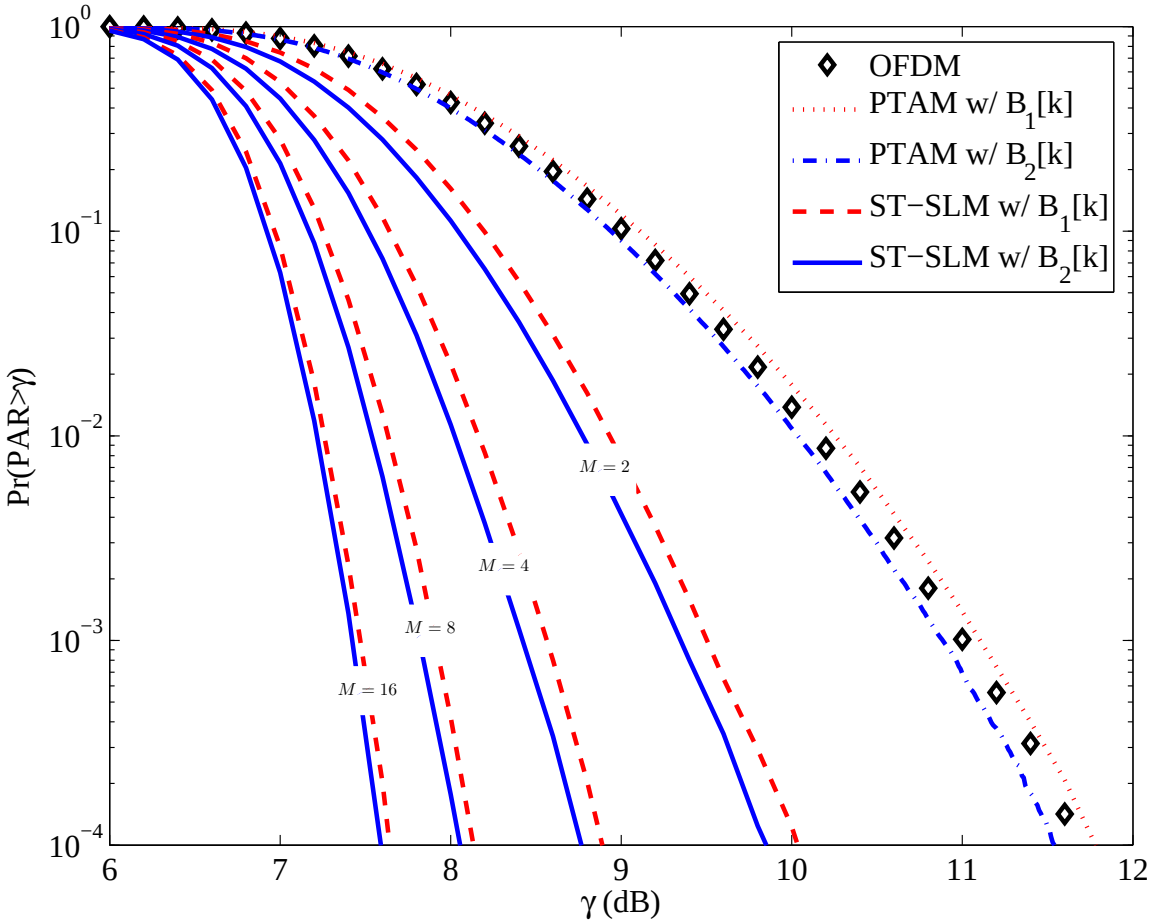


Figure 5.9: CCDF of the PAR in the proposed algorithm. ST stands for superimposed training here.

is used in favor of higher power efficiency. When perfect CSI is available at the receiver, the BER result provides a basis for comparison. The receiver used a zero-forcing equalizer and a suboptimal but simple hard-decision decoder [50]. The frequency selective channel is modeled as Rayleigh fading channels with i.i.d. complex Gaussian taps. The BER was evaluated by averaging over 10^3 Monte Carlo trials.

From Fig. 5.10, we can see that with superimposed training and in the absence of PAR reduction, the simple channel estimate (5.39) had the worst BER performance (solid line with circles). Superimposed training with PAR reduction (SLM with $M = 8$, solid line with dots) had an SNR gain of 2 dB, but the improvement diminished at high SNRs due to the error floor. Superimposed training with the improved channel estimate (5.47) and SLM with $M = 8$ (solid line with squares) lowered the error floor but did not eliminate it. Next, we

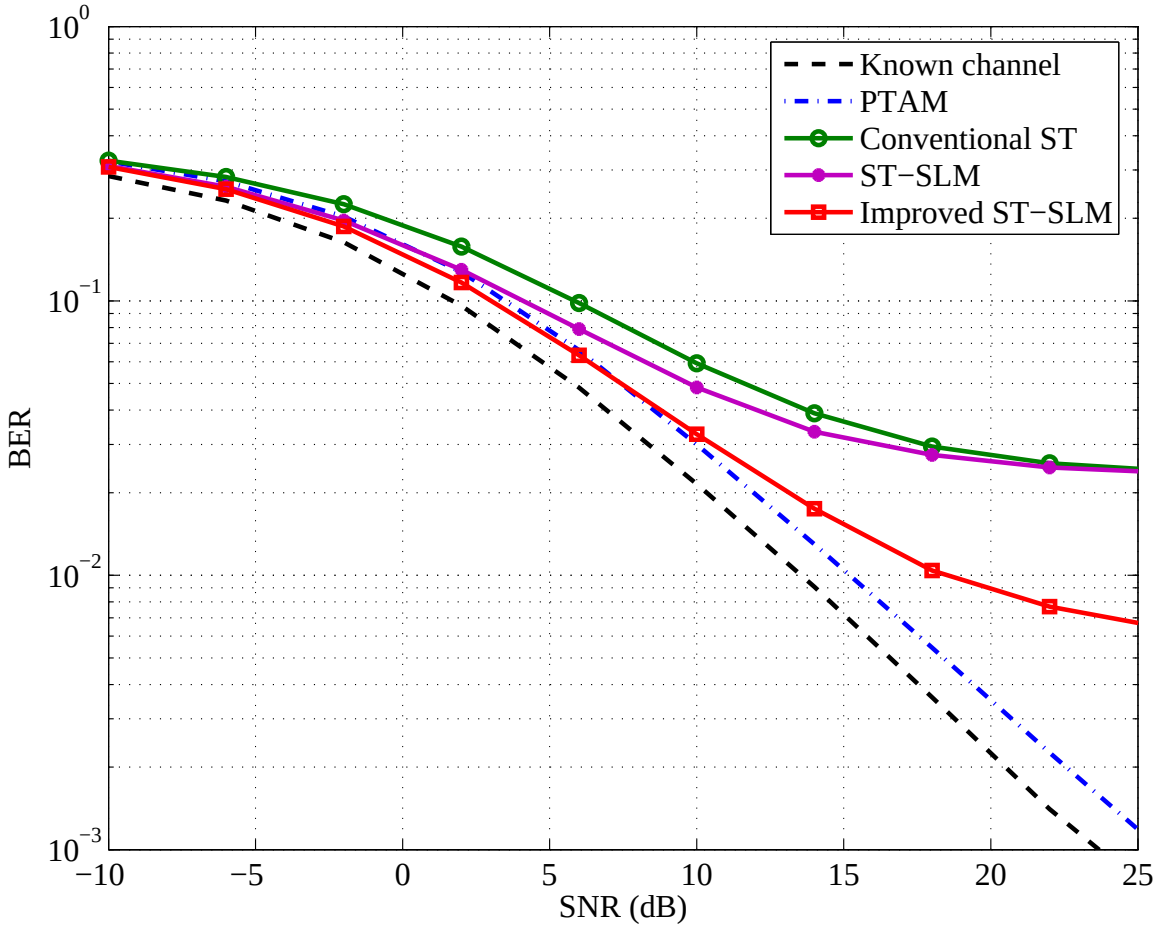


Figure 5.10: BER of the proposed algorithm. ST stands for superimposed training here.

compare our proposed algorithm with PTAM (dash-dotted line) and the known channel case (dashed line). We observe from Fig. 5.10 that PTAM had a very good BER performance as it approaches the known channel case to within $1-2$ dBs. However, superimposed training results in an error floor at high SNRs because the information signal acts as the dominant source of "noise" during channel estimation. On the other hand, at medium to low SNRs (i.e., SNR$<$ 10 dB), our proposed algorithm performs similarly to or better than PTAM, the net benefit being the enhanced bandwidth efficiency.

CHAPTER 6

CONCLUSIONS

This dissertation aims to improve the bandwidth efficiency and power efficiency of wireless transmission systems, such as CDMA and OFDM. To improve the bandwidth efficiency, a superimposed training scheme was proposed to acquire the channel state information: the training pilots are superimposed onto the information data thus saving bandwidth. To improve the power efficiency, we advocate the SLM technique which is a distortionless PAR reduction method that can be applied to both OFDM and forward link CDMA. However, the conventional SLM requires the transmission of side information thus reducing the overall bandwidth efficiency. In this dissertation, we carried out detailed studies on blind SLM (i.e., without explicit side information transmission) techniques for both OFDM and forward link CDMA. Throughout this research, computer simulations and testbed experiments were conducted to demonstrate the performance of the proposed algorithms.

6.1 Contributions

We summarize below primary contributions of this dissertation:

- Proposed a novel superimposed training scheme to estimate doubly selective fading channels.
- Proposed a thresholding and clipping method for the SLM structure to greatly reduce the computational complexity without sacrificing the PAR reducing capability.
- Proposed a blind phase SLM technique and a blind Walsh code SLM technique for forward link CDMA to avoid the transmission of side information.
- Proposed a blind SLM technique called BSPTM to facilitate joint channel estimation and PAR reduction for OFDM.

- Derived CCDF of the PAR of the OFDM signal with superimposed training and determined its achievable lower and upper bounds.

- Proposed a novel blind SLM technique with a generalized superimposed training framework to increase the bandwidth efficiency and power efficiency of an OFDM system.

In addition, for each proposed algorithm, we have carefully considered the bandwidth efficiency - power efficiency - complexity - bit error rate tradeoffs.

6.2 Suggestions for Further Research

The following is a list of interesting research topics that can be pursued as extensions of this dissertation:

- Develop a superimposed training algorithm for the multiple input multiple output (MIMO) system.

- Perform testbed experiments on the proposed blind selected pilot tone modulation (BSPTM) method using a channel emulator.

- Adopt the power efficiency ideas of Section 5.1.3 for the optimization of wireless systems, such as for adaptive modulation and coding, and MIMO systems, where the PAR problem may not have been considered when optimizing the system parameters.

APPENDIX A

PROOFS FOR PROPOSITIONS 5.2 AND 5.3

First, let us define a special function

$$f(\rho) \triangleq \sum_{k=1}^{\infty}\sum_{l=0}^{k-1}\frac{\rho^l}{l!}\frac{\lambda^k}{k!}. \tag{A.1}$$

We infer the following properties for $f(\rho)$:

Property A.1. *The $f(\rho)$ function in (A.1) is logarithmically concave for $\rho \geq 0$; i.e., the 2nd-order derivative of $\ln f(\rho)$ is ≤ 0 for $\rho \geq 0$.*

Proof: See Appendix B.

Property A.2. *The logarithmically concave function $f(\rho)$ satisfies*

$$f(\sum_{n=0}^{P-1}\rho_n)\,[f(0)]^{P-1} \leq \prod_{n=0}^{P-1} f(\rho_n) \leq \left[f(\frac{1}{P}\sum_{n=0}^{P-1}\rho_n)\right]^P. \tag{A.2}$$

Proof: See Appendix C.

Recall the CCDF expression in (5.16). Since we are interested in finding the lower and upper bounds of (5.16) as $p[n]$ varies, we only need to pay attention to the terms on the RHS of (5.16) that are functions of $p[n]$; i.e.,

$$\prod_{n=0}^{P-1}\sum_{k=1}^{\infty}\frac{(\frac{\gamma}{\sigma_s^2/(\sigma_s^2+\sigma_p^2)})^k}{k!}\frac{\Gamma(k,|p[n]|^2/\sigma_s^2)}{(k-1)!}. \tag{A.3}$$

By substituting $\lambda = \frac{\gamma}{\sigma_s^2/(\sigma_s^2+\sigma_p^2)}$, $\rho_n = |p[n]|^2/\sigma_s^2$, and the $\Gamma(k,a)$ expression (5.17) into (A.3), we obtain

$$e^{\sum_{n=0}^{P-1}\rho_n}\prod_{n=0}^{P-1}\sum_{k=1}^{\infty}\sum_{l=0}^{k-1}\frac{\rho_n^l}{l!}\frac{\lambda^k}{k!}. \tag{A.4}$$

Since the exponential term in (A.4), $e^{\sum_{n=0}^{P-1}\rho_n} = e^{P\sigma_p^2/\sigma_s^2}$, does not depend on $p[n]$, bounding (5.16) is equivalent to bounding

$$\prod_{n=0}^{P-1}\sum_{k=1}^{\infty}\sum_{l=0}^{k-1}\frac{\rho_n^l}{l!}\frac{\lambda^k}{k!}. \tag{A.5}$$

Using the definition in (A.1), the quantity in (A.5) is simply $\prod_{n=0}^{P-1} f(\rho_n)$.

Next, we would like to investigate what happens to the inequalities in (A.2) when $p[n]$ is a periodic impulse sequence and when $|p[n]|$ is constant.

First, consider the case when

$$p[n] = \begin{cases} \sqrt{P}\sigma_p, & n = lP,\ l = 0, 1, \ldots, \\ 0, & \text{otherwise.} \end{cases} \tag{A.6}$$

Since $\rho_n = |p[n]|^2/\sigma_s^2$, for the $p[n]$ in (A.6), $\rho_n = 0,\ 1 \leq n \leq P-1$. Thus,

$$f(\sum_{n=0}^{P-1} \rho_n) = f(\rho_0), \tag{A.7}$$

$$\prod_{n=0}^{P-1} f(\rho_n) = f(\rho_0)[f(0)]^{P-1}. \tag{A.8}$$

Combining (A.7) and (A.8), we infer that when $p[n]$ is a periodic impulse sequence,

$$\prod_{n=0}^{P-1} f(\rho_n) = f(\sum_{n=0}^{P-1} \rho_n)[f(0)]^{P-1}. \tag{A.9}$$

From (A.2) and (A.9), we conclude that the lower bound on $\prod_{n=0}^{P-1} f(\rho_n)$ is reached for the $p[n]$ given in (A.6).

Next, let us consider the case when

$$|p[n]| = \sigma_p,\ \forall\ n. \tag{A.10}$$

Since $\rho_n = |p[n]|^2/\sigma_s^2$, $\rho_n = \rho_0,\ \forall\ n$. Therefore,

$$\prod_{n=0}^{P-1} f(\rho_n) = [f(\rho_0)]^P = [f(\frac{1}{P}\sum_{n=0}^{P-1} \rho_n)]^P. \tag{A.11}$$

From (A.2) and (A.11), we infer that the upper bound on $\prod_{n=0}^{P-1} f(\rho_n)$ is reached when $|p[n]|$ is constant.

Consider these results in the context of (5.16) and pay attention to the minus sign in front of the exponential term. We thus prove *Proposition 5.2* and

APPENDIX B

PROOF OF PROPERTY A.1

We prove here that the second-order derivative of $\ln f(\rho)$ is non-positive for any $\rho \geq 0$.

From (A.1), we infer that

$$[\ln f(\rho)]'' = \frac{g(1,\rho)g(3,\rho) - [g(2,\rho)]^2}{[g(1,\rho)]^2}, \tag{B.1}$$

where

$$g(n,\rho) \triangleq \sum_{k=n}^{\infty} \sum_{l=0}^{k-n} \frac{\rho^l}{l!} \frac{\lambda^k}{k!}. \tag{B.2}$$

Furthermore, we can express

$$g(n,\rho) = g(n+1,\rho) + \psi(n,\rho), \tag{B.3}$$

where

$$\psi(n,\rho) \triangleq \sum_{k=n}^{\infty} \frac{\rho^{k-n}}{(k-n)!} \frac{\lambda^k}{k!}. \tag{B.4}$$

Using (B.3), we can express the numerator on the RHS of (B.1) as

$$\begin{aligned} g(1,\rho)g(3,\rho) - [g(2,\rho)]^2 &= g(1,\rho)[g(2,\rho) - \psi(2,\rho)] - [g(2,\rho)]^2 \\ &= g(2,\rho)\psi(1,\rho) - g(1,\rho)\psi(2,\rho). \end{aligned} \tag{B.5}$$

For $\rho \geq 0$, we infer from (B.2) and (B.4) that $g(n,\rho) \geq 0, \psi(n,\rho) \geq 0$. Moreover, eq.(B.3) implies that $g(n,\rho) \geq \psi(n,\rho)$. It is not obvious that (B.5) is ≤ 0. Let us define a more general function

$$q(n,k,\rho) \triangleq g(n,\rho)\psi(k,\rho) - g(k,\rho)\psi(n,\rho). \tag{B.6}$$

The quantity in (B.5) is simply $q(2,1,\rho)$. Our objective next is to prove that $q(n,k,\rho) \leq 0, \forall\ k < n$. We shall achieve this by showing

(i) $$q(n,k,0) = q(n,k,\rho)|_{\rho=0} < 0, \quad \forall\ k < n. \tag{B.7}$$

(ii) $$q'(n,k,\rho) < 0, \quad \forall\ k < n. \tag{B.8}$$

In (B.8), the prime $'$ denotes the derivative w.r.t. ρ. Combing (i) and (ii), we will be able to show that

$$q(n,k,\rho) \leq 0, \quad \forall\ k < n. \tag{B.9}$$

Let us first prove (i). From (B.2), we infer that

$$g(n,0) = g(n,\rho)|_{\rho=0} = \sum_{m=n}^{\infty} \frac{\lambda^m}{m!}. \tag{B.10}$$

From (B.4), we infer that

$$\psi(n,0) = \psi(n,\rho)|_{\rho=0} = \frac{\lambda^n}{n!}. \tag{B.11}$$

The ratio of (B.10) and (B.11) yields

$$\frac{g(n,0)}{\psi(n,0)} = 1 + \sum_{m=1}^{\infty} \frac{\lambda^m}{(n+1)\ldots(n+m)}, \tag{B.12}$$

which increases with decreasing n. Therefore,

$$\frac{g(n,0)}{\psi(n,0)} < \frac{g(k,0)}{\psi(k,0)}, \quad \forall\ k < n. \tag{B.13}$$

Equivalently,

$$q(n,k,0) = g(n,0)\psi(k,0) - g(k,0)\psi(n,0) < 0, \quad \forall\ k < n, \tag{B.14}$$

thus proving (i).

To prove (ii), we need to establish some basic relationships on the derivatives. From (B.2) and (B.4), we infer that

$$g'(n,\rho) = g(n+1,\rho), \tag{B.15}$$

$$\psi'(n,\rho) = \psi(n+1,\rho). \tag{B.16}$$

Substituting (B.15) and (B.16) into (B.6), we obtain

$$\begin{aligned} q'(n,k,\rho) &= g(n+1,\rho)\psi(k,\rho) - g(k,\rho)\psi(n+1,\rho) + g(n,\rho)\psi(k+1,\rho) - g(k+1,\rho)\psi(n,\rho) \\ &= q(n+1,k,\rho) + q(n,k+1,\rho). \end{aligned} \tag{B.17}$$

Since $q(n,n,\rho) = 0$, a special case of (B.17) is

$$q'(n,n-1,\rho) = q(n+1,n-1,\rho). \tag{B.18}$$

From (B.2) and (B.4), we infer that $\lim_{n\to\infty} g(n,\rho) = 0$ and $\lim_{n\to\infty} \psi(n,\rho) = 0$. These together with the fact that $\psi(k,\rho) \leq g(k,\rho)$, imply that there exists a large $n = N$ such that

$$q(N,k,\rho) \leq 0. \tag{B.19}$$

To prove (ii), we follow an inductive approach. We already know that (B.9) holds true for a large N (c.f. (B.19)). We want to establish that it is also true for $n-1$; i.e.,

$$q(n-1,k,\rho) \leq 0, \quad \forall\ k < n-1. \tag{B.20}$$

By the principle of induction, we can thus establish that $q(2,1,\rho) \leq 0$.

Let us assume that

$$q(n,k,\rho) \leq 0, \quad \text{for a given } n,\ \forall\ k < n. \tag{B.21}$$

From (B.18) and (B.21), we find

$$q'(n-1,n-2,\rho) = q(n,n-2,\rho) \leq 0. \tag{B.22}$$

Considering this with (B.7), we deduce that

$$q(n-1,n-2,\rho) \leq 0. \tag{B.23}$$

This is our first step towards (B.20) by showing that (B.20) holds for $k = n-2$. Next, we need to show that (B.20) holds for $k = n-3,\ n-4,\ \ldots$, as well. From (B.17), we have

$$q'(n-1,n-3,\rho) = q(n,n-3,\rho) + q(n-1,n-2,\rho). \tag{B.24}$$

The first term on the RHS of (B.24) is ≤ 0 according to (B.21); the second term on the RHS of (B.24) is ≤ 0 according to (B.23). Therefore, $q'(n-1,n-3,\rho) \leq 0$. Again, considering this with (B.7), we deduce that

$$q(n-1,n-3,\rho) \leq 0. \tag{B.25}$$

We thus prove (B.20) by repeating the above procedure.

In summary, we have established that if $q(n,k,\rho) \leq 0$ for a particular n, $q(n-1,k,\rho) \leq 0$ as well. Since we know that $q(n,k,\rho) \leq 0$, for some large N, by induction, we can go down in n and infer that $q(2,1,\rho) < 0$. This proves that the second-order derivative of $\ln f(\rho)$ in (B.1) is ≤ 0, and $\ln f(\rho)$ is concave.

APPENDIX C

PROOF OF EQ. (A.2)

First recall that any concave function $\phi(\rho)$ satisfies the following inequalities

$$\begin{aligned} \phi(\rho_1+\rho_2)+\phi(0) &\leq \phi(\rho_1)+\phi(\rho_2), & \text{(C.1)} \\ \frac{\sum_{n=0}^{P-1}\phi(\rho_n)}{P} &\leq \phi(\frac{\sum_{n=0}^{P-1}\rho_n}{P}), & \text{(C.2)} \end{aligned}$$

for any $\rho_n \geq 0$.

Since we have established that $\ln f(\rho)$ is concave, we can replace $\phi(\cdot)$ by $\ln f(\cdot)$ in (C.2). Recall that $e^{\ln f(x)} = f(x)$, we thus infer that

$$\left[\prod_{n=0}^{P-1} f(\rho_n)\right]^{\frac{1}{P}} \leq f(\frac{1}{P}\sum_{n=0}^{P-1}\rho_n), \quad \text{(C.3)}$$

or equivalently,

$$\prod_{n=0}^{P-1} f(\rho_n) \leq \left[f(\frac{1}{P}\sum_{n=0}^{P-1}\rho_n)\right]^P. \quad \text{(C.4)}$$

Next, let us examine

$$\phi(\sum_{n=0}^{P-1}\rho_n)+(P-1)\phi(0) = \phi(\rho_0+\sum_{n=1}^{P-1}\rho_n)+\phi(0)+(P-2)\phi(0). \quad \text{(C.5)}$$

Applying (C.1) to the RHS of (C.5), we infer that

$$\begin{aligned} \phi(\sum_{n=0}^{P-1}\rho_n)+(P-1)\phi(0) &\leq \phi(\rho_0)+\phi(\sum_{n=1}^{P-1}\rho_n)+(P-2)\phi(0) & \text{(C.6)} \\ &= \phi(\rho_0)+\phi(\rho_1+\sum_{n=2}^{P-1}\rho_n)+\phi(0)+(P-3)\phi(0). & \text{(C.7)} \end{aligned}$$

Again applying (C.1) to the RHS of (C.7), we infer that

$$\phi(\sum_{n=0}^{P-1}\rho_n)+(P-1)\phi(0) \leq \phi(\rho_0)+\phi(\rho_1)+\phi(\sum_{n=2}^{P-1}\rho_n)+(P-3)\phi(0). \quad \text{(C.8)}$$

From (C.6) and (C.8), we deduce that by repeatedly applying (C.1), we eventually obtain

$$\phi(\sum_{n=0}^{P-1}\rho_n)+(P-1)\phi(0) \leq \sum_{n=0}^{P-1}\phi(\rho_n). \quad \text{(C.9)}$$

Now replace $\phi(\cdot)$ by $\ln f(\cdot)$ and recall that $e^{\ln f(x)} = f(x)$, we infer that (C.9) is equivalent to

$$f(\sum_{n=0}^{P-1} \rho_n)[f(0)]^{P-1} \leq \prod_{n=0}^{P-1} f(\rho_n). \tag{C.10}$$

Combining (C.4) and (C.10), we arrive at

REFERENCES

[1] "Digitally controlled transmitters for next generation communications systems." `http://cwc.ucsd.edu/rp_03-05_I.html`, Date Accessed: Apr. 2005.

[2] "Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03 version 3.2.0)," *ETSI TR 101 112*, April 1998.

[3] Adireddy, S., Tong, L., and Viswanathan, H., "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2338–2353, Aug. 2002.

[4] Aksu, A. and Walton, R., "14.4 kbps forward link performance," *TIA TR45.5 contribution*, Feb. 1995.

[5] Attar, R. A., "On the peak-to-average ratio (PAR) of an IS-856 (CDMA2000 1xEV) forward link," in *Proc. IEEE International Conference on Mobile and Wireless Communication Networks*, (Recife, Brazil), Aug. 2001.

[6] Baüml, R. W., Fischer, R. F. H., and Huber, J., "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *IEE Electronics Letters*, vol. 32, pp. 2056–2057, Oct. 1996.

[7] Braithwaite, R. N., "Exploiting data and code interactions to reduce the power variance for CDMA sequences," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 1061–1069, June 2001.

[8] Breiling, M., Muller-Weinfurtner, S. H., and Huber, J. B., "SLM peak-power reduction without explicit side information," *IEEE Communications Letters*, vol. 5, pp. 239–241, June 2001.

[9] Brillinger, D. R., ed., *Time Series: Data Analysis and Theory*. Philadelphia: Society for Industrial and Applied Mathematics, 2001.

[10] Budianu, C. and Tong, L., "Channel estimation for space-time orthogonal block codes," *IEEE Trans. on Signal Processing*, vol. 50, pp. 2515–2528, Oct. 2002.

[11] Chen, N. and Zhou, G. T., "Superimposed training for OFDM: a peak-to-average power ratio analysis." *IEEE Trans. Signal Processing*, to appear in 2006.

[12] Chen, N. and Zhou, G. T., "A superimposed periodic pilot scheme for semi-blind channel estimation of OFDM systems," in *Proc. 10th IEEE DSP Workshop*, (Pine Mountain, GA), pp. 362–365, Oct. 2002.

[13] Chen, N. and Zhou, G. T., "What is the price paid for superimposed training in ofdm?," in *Proc. IEEE Intl. Conference on Acoustics, Speech, and Signal Processing*, (Montreal, Canada), pp. 421–424, May 2004.

[14] Chen, N. and Zhou, G. T., "Peak-to-average power ratio reduction in OFDM with blind selected pilot tone modulation," in *Proc. IEEE Intl. Conference on Acoustics, Speech, and Signal Processing*, vol. 3, (Philadelphia, PA), pp. 845–848, March 2005.

[15] Chevreuil, A. and Loubaton, P., "Blind second-order identification of FIR channels: Forced cyclo-stationarity and structured subspace method," *IEEE Signal Processing Lett.*, vol. 4, pp. 204–206, July 1997.

[16] Cripps, S. C., *RF Power Amplifiers for Wireless Communications*. Norwood, MA: Artech House, 1999.

[17] Davis, J. A. and Jedwab, J., "Peak-to-mean power control and error correction for OFDM transmission using golay sequences and reed-muller codes," *Elect. Lett.*, vol. 33, pp. 267–268, Feb. 1997.

[18] Ding, L., Zhou, G. T., Morgan, D. R., Ma, Z., Kenney, J. S., Kim, J., and Giardina, C. R., "A robust predistorter constructed using memory polynomials," *IEEE Transaction on Communications*, vol. 52, pp. 159–165, Jan. 2004.

[19] Farhang-Boroujeny, B., "Pilot-based channel identification: proposal for semi-blind identification of communication channels," *Electron. Lett.*, vol. 31, pp. 1044–1046, June 1995.

[20] Farhang-Boroujeny, B., "Experimental study of semi-blind channel identification/equalization through pilot signals," in *Proc. 3rd IEEE Intl. Conf. on Signal Processing*, vol. 1, (Beijing, China), pp. 618–621, Oct. 1996.

[21] García, M. J. F.-G., Edfors, O., and Páez-Borrallo, J. M., "Joint channel estimation and peak-to-average power reduction in coherent OFDM: a novel approach," in *Proc. IEEE Vehicular Technology Conference - Spring*, vol. 2, (Rhodes, Greece), pp. 815–819, May 2001.

[22] Ghogho, M., McLernon, D. C., Alameda-Hernandez, E., and Swami, A., "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Processing Letters*, vol. 12, pp. 226–229, March 2005.

[23] Giannakis, G. B., "Filterbanks for blind channel identification and equalization," *IEEE Signal Processing Lett.*, vol. 4, pp. 184–187, June 1997.

[24] Giannakis, G. B. and Tepedelenlioglu, C., "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. of the IEEE*, vol. 86, pp. 1969–1986, Oct. 1998.

[25] Gradshteyn, I. S. and Ryzhik, I. M., eds., *Table of Integrals, Series, and Products*. Academic Press, 6 ed., 2000.

[26] Hosokawa, S., Teo, K. D., Ohno, S., and Hinamoto, T., "Pilot tone design with low peak-to-average power ratio in ofdm," *IEICE Trans. on Fundamentals*, vol. E88-A, pp. 2117–2123, Aug. 2005.

[27] Jayalath, A., "Adaptive PTS approach for reduction of peak-to-average power ratio of OFDM signal," *Elec. Lett.*, vol. 36, pp. 1226–1228, July 2000.

[28] Jayalath, A. and Tellambura, C., "A blind SLM receiver for PAR-reduced OFDM," in *Proc. IEEE 56th Vehicular Technology Conference*, vol. 1, (Vancouver, Canada), pp. 219–222, Sept. 2002.

[29] Jiang, T. and Zhu, G., "Nonlinear companding transform for reducing peak-to-average power ratio of OFDM signals," *IEEE Trans. Broadcast.*, vol. 50, pp. 342–346, Sept. 2004.

[30] Jones, A. E., Wilkinson, T. A., and Barton, S. K., "Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission scheme," *Elec. Letts.*, vol. 30, pp. 2098–2099, Dec. 1994.

[31] Ju, S. M. and Leung, S. H., "Clipping on COFDM with phase on demand," *IEEE Commun. Lett.*, vol. 7, pp. 49–51, Feb. 2003.

[32] Kastenholz, C. E. and Birkemeier, W. P., "A simultaneous information transfer and channel sounding modulation technique for wide-band channels," *IEEE Trans. on Communication Technology*, pp. 162–165, June 1965.

[33] Kou, Y. J., Lu, W. S., and Antoniou, A., "Peak-to-average power-ratio reduction via channel hopping for downlink CDMA systems," in *Proc. 2003 International Symposium on Circuits and Systems*, vol. 2, (Bangkok, Thailand), pp. 236–239, May 2003.

[34] Krongold, B. S. and Jones, D. L., "PAR reduction in OFDM via active constellation extension," *IEEE Trans. Broadcast.*, vol. 49, pp. 258–268, Sept. 2003.

[35] Lau, V. K. N., "On the analysis of peak-to-average ratio (PAR) for IS95 and CDMA2000," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 2174–2188, Nov. 2000.

[36] Lee, J. S. and Miller, L. E., "Analysis of peak-to-average power ratio for IS-95 and third generation CDMA forward link waveforms," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 1004–1013, July 2001.

[37] Lettieri, P. and Srivastava, M. B., "Advances in wireless terminals," *IEEE Personal Communications*, vol. 6, pp. 6–19, Feb. 1999.

[38] Li, X. and Cimini, L. J., "Effects of clipping and filtering on the performance of OFDM," *IEEE Commun. Letts*, vol. 2, pp. 131–133, May 1998.

[39] Ljung, L., *System Identification: Theory for the User*. Prentice Hall, 1998.

[40] Ma, X., Giannakis, G. B., and Ohno, S., "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. on Signal Processing*, vol. 51, pp. 1351–1366, May 2003.

[41] Manton, J. H., Mareels, I. Y., and Hua, Y., "Affine precoders for reliable communications," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, (Istanbul, Turkey), pp. 2749–2752, June 2000.

[42] Mazzenga, F., "Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence," *IEEE Trans. Broadcast.*, vol. 46, pp. 170–176, June 2000.

[43] Meng, X. and Tugnait, J. K., "Semi-blind channel estimation and detection using superimposed training," in *Proc. IEEE Int. Conf. Acoust, Speech, Signal Processing*, vol. 4, (Montreal, Canada), pp. 417–420, May 2004.

[44] Mestdagh, D. J. G. and Spruyt, P. M. P., "A method to reduce the probability of clipping in DMT-based transceivers," *IEEE Trans. on Communications*, vol. 44, pp. 1234–1238, Oct. 1996.

[45] Müller, S. H. and Huber, J. B., "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequence," *Electron. Lett.*, vol. 33, pp. 368–369, Feb. 1997.

[46] Negi, R. and Cioffi, J., "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electron.*, vol. 44, pp. 1122–1128, Aug. 1998.

[47] Ochiai, H., "Performance analysis of peak power and band-limited OFDM system with linear scaling," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1055–1065, Sept. 2003.

[48] Ochiai, H. and Imai, H., "On the distribution of the peak-to-average power ratio in OFDM signals," *IEEE Trans. Commun.*, vol. 49, pp. 282–289, Feb. 2001.

[49] Ohkubo, N. and Ohtsuki, T., "A peak to average power ratio reduction of multicarrier CDMA using selected mapping," in *Proc. IEEE 56th Vehicular Technology Conference*, vol. 4, (Vancouver, Canada), pp. 2086–2090, Sept. 2002.

[50] Ohno, S. and Giannakis, G. B., "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. Commun.*, vol. 50, pp. 2113–2123, Dec. 2002.

[51] Orozco-Lugo, A. G., Lara, M. M., and McLernon, D. C., "Channel estimation using implicit training," *IEEE Trans. Signal Processing*, vol. 52, pp. 240–254, January 2004.

[52] Proakis, J. G., *Digital Communications*. McGraw-Hill, 3rd ed., 1995.

[53] Qian, H., Xiao, C., Chen, N., and Zhou, G. T., "Dynamic selected mapping for OFDM," in *Proc. IEEE Intl. Conference on Acoustics, Speech, and Signal Processing*, vol. 4, (Philadelphia, PA), pp. 325–328, March 2005.

[54] Rice, S., "Mathematical analysis of random noise," *Bell Systems Technical Journal*, vol. 24, pp. 46–156, Jan. 1945.

[55] Ryu, H. G., Hoa, T. P., Lee, K. M., Kim, S. W., and Park, J. S., "Improvement of power efficiency of HPA by the PAPR reduction and predistortion," *IEEE Trans. Consumer Electronics*, vol. 50, pp. 119–124, Feb. 2004.

[56] Schmidt, R. O., "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Est. Workshop*, (Rome, NY), pp. 243–258.

[57] Serpedin, E. and Giannakis, G. B., "Blind channel identification and equalization with modulation-induced cyclostationarity," *IEEE Trans. Signal Processing*, vol. 46, pp. 1930–1944, July 1998.

[58] Shanbhag, A. G. and Tiedemann, E. G., "Peak-to-average reduction via optimal Walsh code allocation in third generation CDMA systems," in *Proc. IEEE 6th International Symposium on Spread Spectrum Techniques and Applications*, vol. 2, (Parsippany, NJ), pp. 560–564, Sept. 2000.

[59] Slimane, S. B., "Peak-to-average power ratio reduction of OFDM signals using pulse shaping," in *Proc. IEEE GLOBECOM 2000*, vol. 3, (San Francisco, CA), pp. 1412–1416, Nov. 2000.

[60] Sperlich, R., Park, Y., Copeland, G., and Kenney, J. S., "Power amplifier linearization with digital pre-distortion and crest factor reduction," in *Proc. IEEE International Microwave Symposium Digest*, vol. 2, (Fort Worth, Texas), pp. 669–672, June 2004.

[61] Tarohk, V. and Jafarkhani, H., "On the computation and reduction of the peak-to-average power ratio in multicarrier communications," *IEEE Trans. Commun.*, vol. 48, pp. 37–44, Jan. 2000.

[62] Tellado, J., *Multicarrier Modulation With Low PAR – Applications to DSL and Wireless*. New York: Kluwer Academic Publishers, 2000.

[63] Tellambura, C., "Computation of the continuous-time par of an OFDM signal with BPSK subcarriers," *IEEE Commun. Lett.*, vol. 5, pp. 185–187, May 2001.

[64] Triolosd, A. A., "Advanced techniques in high-efficiency power amplification for WCDMA." *NJIT ECE Seminar,* Sept. 2002. `http://www.njit.edu/old/ECE/seminars/seminars.htm`, Date Accessed: Apr. 2005.

[65] Tsatsanis, M. K. and Giannakis, G. B., "Modeling and equalization of rapidly fading channels," *Intl. J. of Adaptive Control and Signal Processing*, vol. 10, pp. 159–176, 1996.

[66] Tsatsanis, M. K. and Giannakis, G. B., "Transmitter induced cyclostationarity for blind channel equalization," *IEEE Trans. Signal Processing*, vol. 45, pp. 1785–1794, July 1997.

[67] Tugnait, J. K. and Luo, W., "On channel estimation using superimposed training and first-order statistics," *IEEE Commun. Lett.*, vol. 7, pp. 413–415, Sept. 2003.

[68] Vääänänen, O., Vankka, J., Viero, T., and Halonen, K., "Reducing the crest factor of a CDMA downlink signal by adding unused channelization codes," *IEEE Commun. Lett.*, vol. 6, pp. 443–445, Oct. 2002.

[69] van Nee, R., ed., *OFDM Wireless Multimedia Communications*. Boston, London: Artech House, 2000.

[70] Wang, C. L., Hsu, M. Y., and Ouyang, Y., "A low-complexity peak-to-average power ratio reduction technique for OFDM systems," *Proc. IEEE GLOBECOM 2003*, vol. 4, pp. 2375–2379, 2003.

[71] Wang, X., Tjhung, T., and Ng, C., "Reduction of peak-to-average power ratio of OFDM system using a companding technique," *IEEE Trans. Broadcast.*, vol. 45, no. 3, pp. 303–307, 1999.

[72] Wei, S., Goeckel, D. L., and Kelly, P. A., "The complex envelope of a bandlimited OFDM signal converges weakly to a Gaussian random process: proof and application." *IEEE Trans. Inform. Theory*, Sept., 2002, submitted. `http://www.ece.lsu.edu/swei`, Date Accessed: Nov. 2004.

[73] Wulich, D., "Peak factor in orthogonal multicarrier modulation with variable levels," *Electron. Lett.*, vol. 32, pp. 1859–1861, September 1996.

[74] Wulich, D., Dinur, N., and Glinowiecki, A., "Level-clipped high-order OFDM," *IEEE Trans. Commun.*, vol. 48, pp. 928–930, June 2000.

[75] Yang, S. C., ed., *CDMA RF System Engineering.* Artech House, 1998.

[76] You, Y., Jeon, W., Paik, J., and Jung, H., "Low-complexity PAR reduction schemes using SLM and PTS approaches for OFDM-CDMA signals," *IEEE Trans. Consumer Electron.*, vol. 49, pp. 284–289, May 2003.

[77] Zhou, G. T. and Chen, N., "Superimposed training for doubly selective channels," in *Proc. IEEE Statistical Signal Processing Workshop*, (St. Louis, MO), pp. 82–85, Sept. 2003.

[78] Zhou, G. T. and Peng, L., "Optimality condition for selected mapping in OFDM." *IEEE Trans. Signal Processing,* to appear in 2006.

[79] Zhou, G. T., Viberg, M., and Mckelvey, T., "Superimposed periodic pilots for blind channel estimation," in *Proc. 35th Asilomar Conference on Signals, Systems, and Computers*, vol. 1, (Pacific Grove, CA), pp. 653–657, November 2001.

[80] Zhou, G. T., Viberg, M., and McKelvey, T., "A first-order statistical method for channel estimation," *IEEE Signal Processing Lett.*, vol. 10, no. 3, pp. 57–60, 2003.

[81] Zhou, G. T., Qian, H., and Chen, N., *Advances in Nonlinear Signal and Image Processing*, ch. Communication System Nonlinearities: Challenges and Some Solutions. Hindawi, 2006.

VITA

Ning Chen was born in Handan, Hebei, China in 1975. He received his dual B.S. degrees in Electronic Engineering and in Accounting from the Shanghai Jiao Tong University (SJTU), China in 1997. He worked as an instructor at SJTU until August 2000. He received his M.S. degree in Electrical and Computer Engineering from the New Mexico State University in December 2001. Since then, he has been working towards his Ph.D. degree in Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta, Georgia, USA.

His general research interests are in the areas of signal processing and communications. Specific current interests include predistortion linearization of nonlinear power amplifiers, peak-to-average power ratio reduction of communication signals, communication channel identification and equalization, and adaptive algorithm development on DSP.